THIS WEEK

EDITORIALS

DATA New journal from Nature Publishing Group goes back to research basics **p.142**

WORLD VIEW How to speed up the personal-genomics revolution **p.143**

WEAPONS GRADE Rhinoceros beetle horn as good as it looks **p.144**

Mind how you go

The protracted battle to find cures for psychiatric illnesses is changing course, but prejudice and stigma against those with poor mental health remain a problem.

When *Nature* broke the news in December 2011 that Novartis was to close its brain research facility in Basel, Switzerland, the story noted that this move closely followed similar cuts by rival pharmaceutical firms. As a report from the European College of Neuropsychopharmacology had warned a few months earlier, the abandonment of research into new psychiatric drugs by AstraZeneca, GlaxoSmithKline and others meant that "withdrawal of research resources is a withdrawal of hope for patients and their families".

That 2011 story did offer a glimmer of optimism. Novartis, we reported, was not deserting these patients and families completely. It planned to switch focus — away from conventional neuroscience-drug development based on chemistry and small molecules, and towards treatments based on the genetics of psychiatric and cognitive disorders.

As we report this week on page 153, the company has made good on its promise. It has reopened its neuroscience division, now at its global drug-discovery headquarters in Cambridge, Massachusetts, and has hired Ricardo Dolmetsch, a former senior director at the Allen Institute for Brain Science in Seattle, Washington, to head it.

In the time that Novartis has been out of the game, some things have changed in the field of psychiatric medicine, but some have not. Among the latter is the depressing ease with which mental illness is subject to stigma and misunderstanding. Just this month, two UK supermarkets — Asda and Tesco — were criticized for selling online 'mental patient' Halloween costumes, complete with bloody cleavers, and orange jump-suits stamped with the phrase 'Psycho Ward'.

Both companies apologized, and removed the offending items from their digital shelves. If that was a clumsy mistake, and one that seemed to cash in on the Hollywood stereotype of the deranged killer rather than a deliberate move to offend, it is harder to find excuses for *The Sun* newspaper. On 7 October, the UK tabloid splashed on its front page that "mental patients" had killed 1,200 people in a decade (2001–10). Drawing on a sober report from the Centre for Mental Health and Risk at the University of Manchester, the newspaper mangled the results and managed to make the opposite point to that intended by the report's authors about help for people with mental illnesses. Contrary to the newspaper's claims, significant numbers of the 1,200 people highlighted were not receiving any treatment, and so were not "patients" and were under no supervision by the "broken" system. Of those who were being treated, many had alcohol and drug problems rather than severe delusional disorders. And, as the report made clear, homicides by patients with schizophrenia in the United Kingdom — the peg for the paper's "exclusive investigation" was the manslaughter conviction last week of a man with schizophrenia — have fallen to an all-time low.

One heartening change seems to be that the prejudice written through *The Sun*'s story now seems out of step with the public mood — or at least with the mood of the public active on social media. Rapid Twitter campaigns saw the costume-selling supermarkets lambasted and, as *Nature* went to press, *The Sun* was facing similar online bombardment.

The science of mental health has changed too, as Novartis knows well. Earlier this year, when the American Psychiatric Association released the fifth edition of its *Diagnostic and Statistical Manual of Mental Disorders*, the official guide to mental illness, critics complained loudly about its listed categories, such as depression and bipolar disorder, and the way they are classified and diagnosed. Various books have poked holes in psychiatry and its relationship with the drug industry. And prominent researchers talk of the need to go back to the drawing board.

> ***"Millions of people are left exposed, vulnerable and ripe for tabloid demonization."***

Some of this is down to the age-old antipathy of psychologists who prefer a holistic approach to mental illness and who dislike what they view as the reductionism of those who look to biochemistry for the fixes, if not always the causes. Some is down to deep distrust of big pharma, and staggering statistics that record the ever-growing diagnoses of mental disorders and prescriptions for drugs to combat them. But much is also down to sheer frustration that decades of research into psychiatric medicine have failed to find cures for these ailments, leaving millions of people exposed, vulnerable and ripe for tabloid demonization.

Novartis is not the only drug giant to adopt a new approach to this problem. Roche is taking a similar path, and others will surely follow, if only because of the size of the market if they get it right. That is far from certain. The science is immature and this approach looks more than a little like a leap of faith. But all who have a mental illness, and all who know someone affected (almost everybody), should wish them luck. ■

Closed question

The US shutdown is damaging science, and Congress must be called to account.

On 1 October, lawmakers in Congress, bitterly divided along partisan lines, failed to agree on a new budget. The US government closed. Roughly 800,000 civil servants, including thousands of scientists, were ordered to stay at home. Even accessing their work e-mail would constitute a federal crime, they were told. Now entering its second week, the shutdown is showing few signs of abating.

Non-government scientists must be imagining that this nightmare will pass. This is just a Washington DC thing, right?

It is true that the effects are most pernicious in greater Washington, where commuter buses and trains have plenty of empty seats. At the National Science Foundation (NSF) in Arlington, Virginia, 98.5% of

the agency's employees were sent home. But at the National Institutes of Health in Bethesda, Maryland, about one-quarter of the agency's 19,000 employees have kept working, keeping mice fed and cell lines growing. And 45% of workers at the National Oceanic and Atmospheric Administration in Washington DC have kept at it, largely because data collected by the National Weather Service are so crucial.

There is a veneer of continuity. But it is an illusion. Clinical trials will not begin. Grant applications will not be evaluated. Even grants that have been awarded are in jeopardy if the cheque was not in the post.

And the consequences are not confined to Washington DC. Websites used routinely in research are not accessible. Conferences in which government scientists have vital roles are either being cancelled or going ahead as pale shadows of what they ought to have been.

The ripple effects will get worse as the government misses lump-sum payments to contractors. One casualty came on 4 October, when the National Radio Astronomy Observatory, based in Charlottesville, Virginia, closed its radio telescopes. This week, an NSF contractor is preparing to evacuate research stations in Antarctica, putting an entire summer season of research under threat. By the end of the month, the operator of NASA's most famous observatory — the Hubble Space Telescope — will nearly have run out of money (see go.nature.com/smgwr1).

The public data collected at these facilities are used widely. Even scientists at financially secure institutions will soon find themselves missing a key tool or piece of data. In an interconnected world that relies on global collaboration, foreign scientists are not safe either.

Part of the frustration with the fiscal crisis is how manufactured it is. Republicans in the House of Representatives want to extract concessions on US President Barack Obama's signature health-care law. Obama and the Democrat majority in the US Senate say that they will not be held hostage to these demands.

There are signs that the stand-off will persist until 17 October, when a new forcing factor would kick in: the government exceeding the amount of money that it is legally allowed to borrow. The economic consequences of not raising the debt limit are expected to be immediate and catastrophic, and so the warring parties in Congress might finally be forced to compromise.

That would be welcome. The damage being done to science — the slow business of meticulous data gathering — is not as immediately apparent as in other arenas. But it is insidious. A missed moment in a data campaign may not reveal its importance until much later. A talented scientist, fed up with budget vagaries, might seek greener pastures. And an experiment not performed might seem to be no worse than an unasked question. For these reasons, we must all ask the US Congress: why are you doing this? ■

ANNOUNCEMENT

Launch of an online data journal

Everyone wants better ways to make research data available and to give more credit to the researchers who create and share data. But for a data set to be widely reusable, scientists need to know how the data were produced and what quality-control experiments were performed. They need access to detailed descriptions of the data outputs, file formats, sample identifiers and replication structure. This is hard work that is often poorly rewarded. As a result, potentially valuable data sets go unpublished, or are not fully released to the public or not described in sufficient detail to permit reuse.

To address this need, Nature Publishing Group will next spring launch *Scientific Data*, an open-access, online-only journal for detailed descriptions of data sets (http://nature.com/scientificdata). This week, *Scientific Data* announced its first call for submissions (see go.nature.com/1gnd1j). The doors are now open for scientists to submit 'Data Descriptor' manuscripts — a new article type that is designed to describe scientifically valuable data sets in a way that will promote data sharing and reuse.

Data Descriptor articles are fully fledged, peer-reviewed scientific publications, and will be listed in major indexing services, thereby giving authors the credit they deserve for sharing their data and making it usable by others. All Data Descriptors will be released under a Creative Commons licence that allows researchers to reuse, redistribute and remix the articles' content.

The format of the Data Descriptor includes 'Technical Validation' and 'Usage Notes' sections. These will allow authors to characterize the quality of the data and to provide advice on their reuse — valuable information that does not always fit into traditional research articles. And, as is the case in other *Nature* journals, the Methods section will have no length limit, giving authors space to provide detailed, reproducible descriptions of their experiments.

Data Descriptors will link to both related journal articles and data files stored at data repositories, helping readers to navigate easily between research, data descriptions and the actual data. And each Data Descriptor publication will be supported by machine-readable experimental metadata to help advanced users mine and search *Scientific Data*'s content. Metadata records will be curated by in-house staff to ensure consistent and useful annotation, and will be released in the ISA-Tab format (see S.-A. Sansone *et al. Nature Genet.* **44,** 121–126; 2012).

Peer reviewers of Data Descriptors will focus on the technical rigour of the data-collection procedures, the completeness of the data and alignment with existing community standards. They will check that the data are indeed worth sharing, but will specifically be asked not to base their evaluations on the perceived impact or novelty of the findings associated with the data sets. *Scientific Data*'s editors have already conducted peer review of a small set of prototype Data Descriptor manuscripts, and have found that scientists adapt quickly to this different peer-review perspective.

What *Scientific Data* will not be is a new data repository. Rather, it will promote and cooperate with existing community-based repositories, and will combat data fragmentation by ensuring that data sets are deposited in an appropriate repository. *Scientific Data* is also working with figshare and Dryad, two repositories that accept a wide range of research data types. Integrated data upload is already available with figshare — authors may deposit their data as they submit their Data Descriptor manuscript. Editors and referees will be given secure, confidential access to the data files through the figshare website, and the data will be made public when the Data Descriptor is published.

Scientific Data will not be a place to publish new conclusions or hypothesis-driven analyses, and editors will ask authors to remove material that is beyond the journal's scope. This will help to ensure that Data Descriptor publications can exist alongside and complement primary research articles. Authors may publish standalone Data Descriptors about data sets that have not been used in other publications, or Data Descriptors about data sets published elsewhere but for which a more in-depth description is merited.

Editors of *Nature* journals have agreed that prior publication of a Data Descriptor will not jeopardize publication of research articles, as long as those articles go beyond a descriptive analysis of the data and report major scientific findings. *Scientific Data* will initially focus on the life, biomedical and environmental sciences, but may in due course be open to a broader range of scientific disciplines. ■

VOLKER STEGER/SPL

Improving genome understanding

The cost and accuracy of genome sequencing have improved dramatically. **George Church** *asks why so few people are opting to inspect their genome.*

Readers of *Nature*, we can assume, are bright and insatiably curious. So why have so few obtained and interpreted their own genome sequence? We should avoid being judgemental of people who practise genomic modesty or who choose not to act on genome information, but we should also ask if we are providing adequate and equal access to education about the benefits and risks of genome information.

For 7 years I led one of the teams registered to compete for the US$10-million Archon Genomics X Prize, and I was naturally disappointed by the abrupt cancellation of the competition in August. However, the confusion surrounding the X Prize does provide an occasion to reflect on the problems and misunderstandings in genomics. The first is that genomics is seen as expensive. In fact, sequencing costs have plummeted — from $2.7 billion for the first human genome in 2003 down to $1,000 today. That's not much more than the cost of a decent laptop, and much less than a car. However, people are reluctant to pay to have their genome sequenced — many feel that health care should be provided for free by insurance or the government and, indeed, this is our not-that-distant goal, as there are many in our community who would not benefit from genome information if it were not free. However, for those today who can afford a genome sequence, we would argue that, overall, the cost of sequencing is expected to be recovered over a lifetime through the avoidance of unnecessary diagnostics and therapeutics and time spent in waiting rooms and hospitals.

Perhaps too many think that genomics is inaccurate. When it announced the cancellation, the X Prize Foundation claimed that "no company is sequencing whole genomes to the accuracy the contest required". Aside from the pre-judgemental weirdness, is this statement true? Haplotype phasing quality — a measure of accuracy — has improved from 350 kilobases in 2007 to 2,463 kilobases in 2013, and point errors have improved from 1 in 100,000 to 1 in 10 million — both well beyond the X Prize goals. Genetic analyses of tough tandem repeats are now common diagnostically.

Are the results uninterpretable? Even if we place the As, Cs, Gs and Ts in the right order, how does this help? Genome-wide association studies (GWAS) and studies of twins can give the impression that predicting traits from genomic sequence is a haphazard science. But since 1991 the number of highly predictive gene tests has risen from two to 3,000. Even 'complex' traits include components that can be identified and applied clinically to individuals who are not classed to be directly at risk. For example, height and diabetes GWAS have shown that a vast number of common variants have small effects, but the alternative of seeking rare variants reveals large effects by altering levels of growth hormone for height and insulin for diabetes. These hormones are effective therapies even for individuals who are not mutant in them. Too often the messy results of GWAS and twin studies are down to poor selection of subjects and neglect of confounding environmental factors.

➲ NATURE.COM
Discuss this article online at:
go.nature.com/li69jt

Even if they are interpretable, are the results useful? Yes! Even if there is no cure for the genetic conditions identified, there are effective preconception and prenatal options that could have an impact on the family. For example, Ashkenazi communities already use genetic screening to make lists of suitable marital partners early in life to avoid their offspring developing painful Tay–Sachs disease and more than 20 similarly devastating diseases (which are not restricted to their community, by the way). Although we are tempted to restrict genomics to those with ethnic or family risks, the fact is that we are all at risk. Even the possibility of finding markers for one treatable disease (such as a cancer or cardiomyopathy) could, for some, be a sufficient reason to check one's genome.

THOSE OF US AT THE SHARP END OF GENOMICS NEED TO WORK EQUALLY HARD AT CONVERSATIONS WITH THE PUBLIC.

Perhaps most provocatively, some critics assert that genomics could be harmful. The US Genetic Information Nondiscrimination Act (GINA) prevents discrimination based on health insurance and employment; however, there is not a GINA in every country, and it doesn't cover the military, life insurance or person-to-person discrimination. But the question is: do the overall benefits of genomics exceed the risks? Do the benefits of driving trump the one-and-a-quarter million traffic-related deaths per year? A growing number of bioethicists and researchers are worried that typical consenting practices do not inform patients of the likelihood of data escape and re-identification. Certainly, conventional consents served to protect the researchers, not the volunteers. However, the huge numbers of volunteers who are willing to share their genetic data make this a moot point. Why insist on recruiting those — and setting policy around those — who would be upset if their data escapes?

It is important for those of us at the sharp end of work on genomics to work equally hard at conversations with the public. We already share our (very revealing) faces, voices and opinions. And, as we share more of our genetics and as we develop genomic progress into precision medicine, researchers and the public alike need frank assessments of all of these tests and treatments. We need the Genomics X Prize more than ever. ■

George Church *is professor of genetics at Harvard Medical School, Boston, Massachusetts, and founder of the Personal Genome Project.*
e-mail: gmc@harvard.edu

RESEARCH HIGHLIGHTS

Selections from the scientific literature

ASTRONOMY

Cosmic jets trigger star birth

Stars could be forming in an unusual way in a distant galaxy — in the backwash of a massive jet of gases spraying from the galaxy's core.

A team led by Yasir Rashed of the University of Cologne, Germany, used the MERLIN radio telescope array in the United Kingdom to study a galaxy some 1.4 billion parsecs from Earth. Gases shoot outwards from the galaxy's centre, presumably fuelled by a supermassive black hole.

Further observations showed that stars are probably forming near the galaxy. The authors suggest that the outflowing jets produce pressure pockets in the gas outside the galaxy, creating ideal spots for star birth.

Astron. Astrophys. 558, **A5 (2013)**

CYRIL RUOSO/JH EDITORIAL/MINDEN PICTURES/GETTY

ZOOLOGY

Seabird stress response is oceans apart

Seabirds that lay more eggs and die young are more likely to look after their chicks in times of stress than are longer-lived, less-fertile birds.

Jannik Schultner of the Norwegian University of Science and Technology in Trondheim and his colleagues implanted Atlantic and Pacific populations of black-legged kittiwakes (*Rissa tridactyla*) with tubes of corticosterone, a hormone associated with food shortages.

Compared with Atlantic kittiwakes, Pacific populations of the bird have fewer offspring and higher adult survival rates. Pacific kittiwakes with artificially boosted hormones neglected their young, so their chicks were more likely to die than Atlantic ones. Extra hormones had the opposite effect on the Atlantic colony, and actually increased chick survival rates.

Animals' life strategies can predict their short-term responses to stress, the authors suggest.

Proc. R. Soc. B 280, **20132090 (2013)**

EVOLUTION

Big weapons have little downside

The huge horns on male rhinoceros beetles come surprisingly cheap.

Biologists have long thought that the extreme body parts that males use to gain mates — such as showy plumage or aggressive antlers — are costly because they make the animals vulnerable in other ways. By that logic, the horns of a male rhinoceros beetle should exact a toll.

But when Erin McCullough and Douglas Emlen at the University of Montana in Missoula set about quantifying probable costs in dozens of wild and laboratory-raised beetles (*Trypoxylus dichotomus*; **pictured**), they found none. Horn size had little or no effect on flying ability, survival, immune response or the growth of other body structures.

The researchers suggest that weapons may evolve along with traits that let males make best use of them, offsetting survival costs. The low cost could help to explain how the beetles' horns became so elaborate and diverse, the authors say.

Anim. Behav. **http://doi.org/n34 (2013)**

DOUG EMLEN

PERSONAL GENOMICS

A tool to interpret tricky mutations

Researchers have developed software to predict whether certain genetic variants are harmful.

The effects of most mutations are unclear, especially for those in the 99% of the genome that does not code for proteins. Chris Tyler-Smith at the Sanger Institute in Hinxton, UK, and Mark Gerstein at Yale University in New Haven, Connecticut, and their colleagues took non-coding regions that had been identified as functional in a large-scale genomics project and used sequencing data from more than 1,000 people to catalogue how these regions varied in healthy individuals.

This revealed likely patterns of harmful mutations, such as those in DNA sequences to which regulatory proteins bind. The scientists incorporated the patterns into a predictive tool and applied it to genomes from cancer biopsies. This found nearly 100 non-coding variants that could contribute to the disease.

Science 342, **84 (2013)**

SHILONG YANG

MATERIALS SCIENCE

Folded batteries pack power

Flexible lithium-ion batteries folded using origami techniques can store energy in a smaller area. Candace Chan and her colleagues at Arizona State University in Tempe built batteries on paper by attaching lithium metal oxide electrodes to paper coated with conductive carbon nanotubes.

They folded the paper into layers using the Miura-ori pattern, which has also been used to fold solar panels in space. Energy stored per unit area increased by a factor of 14 over the unfolded paper. Because paper-based batteries do not perform as well as standard lithium-ion batteries, the per-area capacity of the folded battery still falls below that of existing batteries.

The researchers think that other folding algorithms and robot-assisted folding would improve performance.

Nano Lett. **http://doi.org/n4d (2013)**

NEUROSCIENCE

Centipede toxin kills pain

A peptide isolated from centipede venom is as potent as morphine at treating pain in mice.

Glenn King of the University of Queensland in St Lucia, Australia, and his colleagues purified a molecule made of 46 amino acids from the venom of the Chinese red-headed centipede (*Scolopendra subspinipes mutilans*, **pictured**). When tested in rat neurons *in vitro*, the peptide strongly inhibited a pain-associated sodium ion channel. It had little effect on related channels. The researchers also injected the molecule into mice that were then exposed to noxious chemicals or heat: the higher the dose, the less the animals reacted to painful stimuli. They showed no obvious side effects. The researchers suggest that this molecule, and perhaps others from centipede venoms, could yield powerful pain treatments.

Proc. Natl Acad. Sci. USA **http://doi.org/n35 (2013)**

REGENERATIVE MEDICINE

Mouth-watering artificial glands

Bioengineered glands that make tears or saliva can take on the function of their natural counterparts when transplanted into mice.

Takashi Tsuji and his colleagues at the Tokyo University of Science harvested cells destined to form tear and salivary glands from mouse embryos and cultured them for three days in plastic devices that mimic conditions in developing embryos.

The bioengineered glands were implanted in mice from which the natural glands had been removed. The tissue then matured and connected with the animals' nervous systems and ducts for tears and saliva, producing the fluids in response to appropriate stimuli.

Nature Commun. 4, **2497; 2498 (2013)**

PHYSIOLOGY

Hormones tick on epigenetic clock

A mammal's annual reproductive cycle is controlled by reversible changes to DNA.

Siberian hamsters (*Phodopus sungorus*) breed during the longest days of the year; shorter days trigger a series of hormonal changes that cause their gonads to shrink. The change is tied to the nocturnally produced hormone melatonin, but how the hormone alters gene activity has been unclear.

Tyler Stevenson, now at the University of Aberdeen, UK, and Brian Prendergast of the University of Chicago in Illinois found that both shorter days and corresponding changes in melatonin reduce methylation of a key gene in the hypothalamus to launch the gonad-shrinking cascade.

The study is the first to show that cycles of DNA modification underlie seasonal changes, the authors say.

Proc. Natl Acad. Sci. USA **http://doi.org/nz8 (2013)**

COMMUNITY CHOICE

The most viewed papers in science

MATERIALS SCIENCE

Dirty water gets squeezed clean

HIGHLY READ on pubs.acs.org in September

Gels stuffed with nanoparticles could offer a quick, effective way to disinfect drinking water after disasters.

Silver nanoparticles are powerful microbicides, but they tend to clump, limiting contact between bacterial cells and the killing surface. Xiao Hu, Teik-Thye Lim and their colleagues at Nanyang Technological University in Singapore uniformly embedded the nanoparticles throughout a sophisticated, super-absorbent sponge.

They used the flexible, lightweight material to soak up bacterium-contaminated water for 15 seconds, then simply squeezed the water out. Microbe numbers in the recovered water fell by as much as ten million times. The gel can be repeatedly reused and, the researchers suggest, could be developed into a pocket-sized water filter for emergencies.

Environ. Sci. Technol. 47, **9363–9371 (2013)**

MICROBIOLOGY

New bacteria from the bowels

A new evolutionary branch of bacteria has been discovered in human guts and in groundwater.

While characterizing microbial communities in human faeces and environmental samples, a team led by Jillian Banfield at the University of California, Berkeley, and Ruth Ley at Cornell University in Ithaca, New York, found DNA sequences that were very distantly related to photosynthetic cyanobacteria. Further sequencing produced three complete and three partial genomes for the bacteria. The authors assigned the microbes their own phylum — Melainabacteria.

The melainabacteria lack the genes required for photosynthesis and for coping with oxygen. They probably generate energy by fermenting various carbon-containing molecules, and may provide their human hosts with the vitamins B and K.

eLife 2, **e01102 (2013)**

CORRECTION

The Research Highlight 'Sound shakes semiconductors' (*Nature* **502,** 8–9; 2013) described methods to improve a film's superconductivity. It should have read 'conductivity'.

SEVEN DAYS *The news in brief*

RESEARCH

ALMA milestone

Less than one month after resolving a major labour dispute (see *Nature* http://doi.org/n37; 2013), the Atacama Large Millimeter/submillimeter Array (ALMA) observatory in Chile announced the completion of its antenna-construction phase on 1 October, when the last of 66 research radio antennas was delivered to the site. North American partners in the US$1.4-billion project have provided 25 antennas; European collaborators have delivered 25; and Asian partners have contributed 16. By the end of the year, all of the antennas are expected to be working together as a single telescope.

Comet flyby

Astronomers got a sneak peek at the much-anticipated comet ISON on 1 October, when it passed within 11 million kilometres of Mars. The Mars Reconnaissance Orbiter snapped images of the comet, which will fly within 1.2 million kilometres of the Sun in November. Not all ISON studies have gone so well: a NASA balloon-borne mission to study the comet experienced a problem after its launch on 28 September and could not collect crucial data.

HIV-antibody trials

The US National Institute of Allergy and Infectious Diseases last week began the first human trials of an antibody to block infection by most strains of HIV. Three HIV-infected adults were infused with the VRC01 antibody, said principal investigator Barney Graham on 3 October. The study will eventually test the antibody's safety in at least 15 people, but no more participants can be enrolled until the US government reopens. If proved safe, VRC01 might be tested on babies in the United States and Africa (see *Nature* **493,** 279–280; 2013).

STAN CHURCHES, NOAA FISHERIES/AP

Walruses seek shelter on land

Roughly 10,000 walruses have come ashore in Alaska (pictured) because of a lack of sea ice, the Associated Press reported on 1 October. The animals began gathering on a barrier-island beach near the village of Point Lay in mid-September, according to aerial surveys conducted by the US National Oceanic and Atmospheric Administration. Female walruses and their calves normally spend the summer and early autumn in the Chukchi Sea, drifting north on sea ice that gives them a platform from which to dive for clams, worms and snails. But a lack of ice this year forced them to shelter on land, placing the skittish animals at risk of stampede as they crowd together. In 2009, 130 young walruses died in a stampede near Icy Cape, Alaska.

Journal sting

The outcome of a 'sting' operation described in *Science* last week suggests that peer review is inadequate or non-existent at many open-access journals that charge publishing fees. Journalist John Bohannon sent concocted and obviously flawed papers to 304 journals, 157 of which accepted the papers for publication. The Open Access Scholarly Publishers Association, based in Utrecht, the Netherlands, said in a response that the report was useful but limited, as Bohannon did not select the journals at random or include subscription-based publications or open-access journals that do not charge publishing fees as controls.

POLICY

Marine threats

The world's oceans face greater danger from human activity than is generally recognized, said marine scientists from the International Programme on the State of the Ocean on 3 October. The group warns that the severity of pollution, overfishing and climate change is being ignored or going unnoticed, and that risks to marine ecosystems have been "significantly underestimated" in previous analyses. The group, backed by the International Union for Conservation of Nature, calls for community management of marine resources, international governance of the high seas and reductions in carbon emissions.

AWARDS

Nobel prizes

Neuroscientist Thomas Südhof and cell biologists James Rothman and Randy Schekman have won this year's Nobel Prize in Physiology or Medicine for their discoveries of mechanisms that regulate vesicles — structures that transport molecules within and between cells. The physics prize went to François Englert

and Peter Higgs for their discovery of what is popularly known as the Higgs boson. See pages 149 and 150 for more. *Nature* went to press before the chemistry prize was awarded, but full details will be available at go.nature.com/a4gvbl.

NATIONAL FEDERATION OF THE BLIND

PEOPLE

UNESCO head

The United Nations Educational, Scientific and Cultural Organization (UNESCO) on 4 October elected Irina Bokova for a second term as director-general. The nomination of Bokova, formerly Bulgaria's ambassador to France, is expected to be approved by UNESCO's general conference in November. Bokova has seen UNESCO through tough financial times since the withdrawal in 2011 of US funding — then 22% of the agency's budget — following the organization's decision to admit Palestine as a member (see *Nature* **479**, 159; 2011).

Braille coder dies

Abraham Nemeth (**pictured**), the blind designer of an internationally recognized Braille code for mathematics, died on 2 October, reported the American Foundation for the Blind in New York. He was 94. For his work on the Nemeth Code of Braille Mathematics and a lifetime of

research on and development of other Braille codes, Nemeth received the Louis Braille Award from the International Braille Research Center in Baltimore, Maryland, in 2006. He was a long-time mathematics professor at the University of Detroit Mercy in Michigan.

BUSINESS

Farming forecasts

Agriculture giant Monsanto announced on 2 October that it has bought a crop-insurance company for about US$930 million. The Climate Corporation, based in San Francisco, California, crunches data from remote-sensing operations, including 2.5 million local weather-monitoring stations around the United States, to assess the risk to farmers from extreme weather events. Its big-data expertise is expected to help Monsanto, which is based in St Louis, Missouri, to improve predictions of how climate change will affect genetically modified crops.

Genetics patent

Genetics firm 23andMe announced on 1 October that it had been awarded a US patent for a method to use parental DNA to predict a baby's traits, including eye colour and disease risk. The patent application, filed more than five years ago, also covers use of the technique to screen sperm and ova for *in vitro* fertilization. Bioethicists expressed concern that the patented method could be used to produce 'designer babies', but the company, which is based in Mountain View, California, says that it does not intend to pursue that application of its technology. See go.nature.com/ruvkrh for more.

Merck staff cuts

Merck, a pharmaceutical powerhouse headquartered in Whitehouse Station, New Jersey, announced on 1 October that it aims to save US$2.5 billion by the end of 2015 by laying off 20% of its employees. About half of the spending cuts will be in research and development. Merck, like many pharmaceutical firms making such moves, says that it plans to bolster alliances with outside companies.

COMING UP

12–17 OCTOBER
Highlights at the annual meeting of the American Society for Reproductive Medicine in Boston, Massachusetts, include research on male contraception, the genetics of the uterine condition endometriosis and developments in egg-freezing.
go.nature.com/vtpab1

EVENTS

Fukushima woes

An international team will review remediation efforts later this month in areas affected by the Fukushima nuclear accident, the International Atomic Energy Agency announced on 4 October. The power plant has suffered several setbacks, including a major radioactive water leak in August (see http://doi.org/n53 and *Nature* **501,** 5–6; 2013) and a smaller spill reported last week. Speaking on 6 October, Japan's prime minister, Shinzo Abe, acknowledged the need for international involvement in containing the leaks.

University rankings

The California Institute of Technology in Pasadena has topped an annual list of the world's best universities, for the third consecutive year. Many European institutions' rankings dropped from previous years, whereas the leading universities in Japan, Singapore, South Korea and China made advances. The list, released on 2 October, is published by the London-based magazine *Times Higher Education*, using information from data company Thomson Reuters to score categories such as teaching, research and citations.

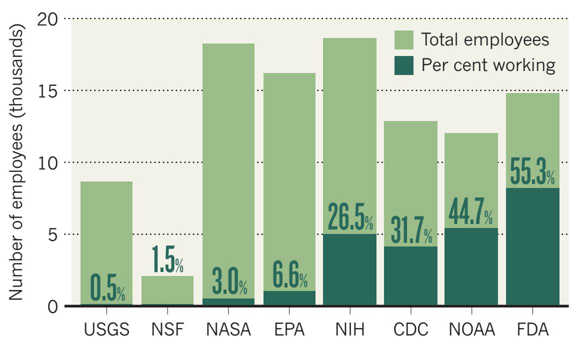
SOURCE: US OMB

TREND WATCH

US government scientists have been hobbled by the government shutdown that began on 1 October, when lawmakers failed to agree on a spending plan for 2014. Science agencies sent most researchers home last week (see chart), and suspended their use of government telephone and e-mail services. Federal scientists were also unable to travel to conferences they had previously planned to attend. Select research operations are being maintained by a minimal set of employees (see *Nature* **502,** 13–14; 2013).

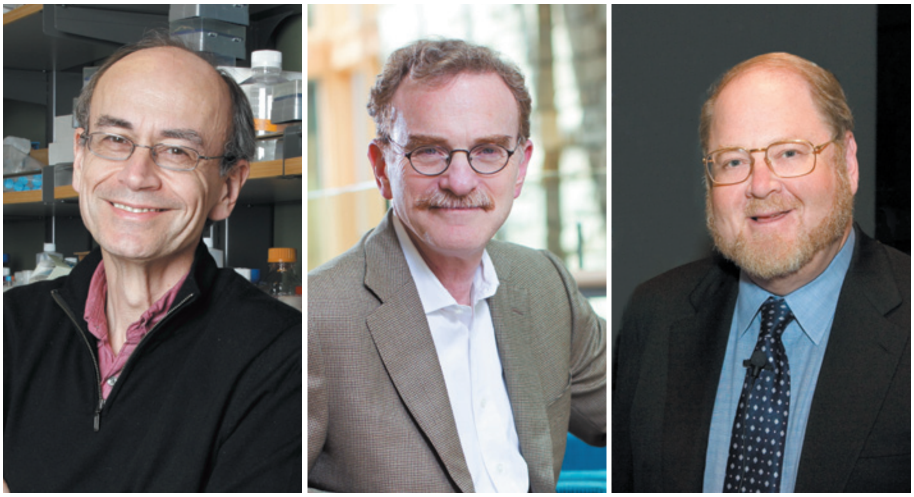
US SHUTDOWN SENDS SCIENTISTS HOME

Government agencies have cut all but 'essential' workers, such as those involved in protecting life or property.



USGS: US Geological Survey; NSF: National Science Foundation; NASA: National Aeronautics and Space Administration; EPA: Environmental Protection Agency; NIH: National Institutes of Health, CDC: Centers for Disease Control and Prevention; NOAA: National Oceanic and Atmospheric Administration; FDA: Food and Drug Administration

➲ NATURE.COM
For daily news updates see:
www.nature.com/news



L TO R: TONY AVELAR; HADAR GOREN; YALE UNIV.

Thomas Südhof (left), Randy Schekman and James Rothman revealed how vesicles interact with targets.

NOBEL PRIZE

Cell transport carries off Nobel

Medicine prize goes to discoverers of vesicle system that shuttles biomolecules around cells.

BY EWEN CALLAWAY

Three scientists who explained the inner workings of a 'cellular postal service' shared this year's Nobel Prize in Physiology or Medicine. Their work pinpointed how cells shuttle proteins and other biomolecules from one location to another — a process that is important in the release of neurotransmitter chemicals, the secretion of insulin and countless other duties.

James Rothman of Yale University in New Haven, Connecticut; Randy Schekman at the University of California, Berkeley; and Thomas Südhof of Stanford University in California, will split the prize — worth 8 million Swedish kronor (US$1.2 million) — which will be awarded at a ceremony on 10 December in Stockholm.

"My first reaction was, 'Oh my God!'" Schekman says. "That was also my second reaction."

The Nobel committee recognized the three researchers "for their discoveries of machinery regulating vesicle traffic, a major transport system in our cells". Vesicles are tiny packages, enclosed by a layer of lipids, that ferry biomolecules around the cell by fusing with other structures.

Hidde Ploegh, an immunologist at the Whitehead Institute in Cambridge, Massachusetts, says that the trio's work is so fundamental to cell biology that it is easy to take it for granted. "When we teach cell biology to biology majors and graduate students, many aspects of vesicular transport are presented as if they've been there all along."

Rothman and Schekman began their work in the late 1970s, when scientists had recognized that vesicles were involved in cellular trafficking, but they knew little about how these structures helped to tame the inner chaos of the cell. To determine how cells make vesicles, how these packages know where to go and how they fuse with other membranes, the two took very different approaches.

Rothman, known for his methodical approach to biochemistry (and noted by some at conferences for his uncanny resemblance to King Henry VIII of England), began as a young researcher at Stanford with the goal of recreating the whole process in a test tube. The work involved looking for proteins by purifying extracts from mammalian cells, some of which had been infected with a virus that, like many other viruses, makes use of the vesicular transport system. One of the viral proteins was used as a marker to track progress through the vesicle system.

Rothman's team purified several proteins that were crucial for vesicles to fuse with their membrane targets. The first was named *N*-ethylmaleimide-sensitive factor, or NSF.

Across San Francisco Bay in Berkeley, Schekman and his team were taking a more diverse approach to studying the same problem. The researchers mutated strains of the yeast *Saccharomyces cerevisiae*, and screened for the ones that could no longer shuttle certain enzymes around the cell. The work eventually turned up 23 'secretory genes', one of which turned out to encode the protein NSF.

Many scientists were initially sceptical of the findings, says Marino Zerial, a molecular biologist who studies cellular trafficking at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany. "The biochemists didn't like the genetic approach and geneticists were doubting the biochemical approach," he says.

But in subsequent work in the late 1980s and early 1990s, the laboratories of Schekman and Rothman worked out the basic mechanisms of vesicle fusion, in which proteins embedded in the membranes of vesicles recognize proteins on the membranes of their destination through a lock-and-key mechanism that is still being worked out. "The synergy that came out from these two is unbelievable," Zerial says. "They're my heroes."

Südhof's work, meanwhile, focused on vesicle fusion in neurons, which communicate with other neurons through neurotransmitter chemicals that are packaged in vesicles. Beginning in the late 1980s, Südhof and a team led by Richard Scheller — now at Genentech, a biotechnology company in ▶

South San Francisco, California — identified a number of proteins that are important to the process. Südhof also showed how the fusion of neurotransmitter-containing vesicles is triggered by calcium.

"It's a fantastic story — it's how in a relatively short time frame, 25 or 30 years, we've gone from knowing almost nothing about how membrane trafficking occurs in cells to knowing the proteins and actors and having a reasonable idea of how they act," says William Wickner, a biochemist at Dartmouth Medical School in Hanover, New Hampshire. "These three are responsible for major, major advances."

Johannes Herrmann, a cell biologist at the Technical University of Kaiserslautern in Germany, was trained by Schekman and had nothing but praise for his former mentor. "It's a great day," he says. He remembers that Schekman would set grand goals for his laboratory and, when they were reached, reward the entire lab with banquet dinners. Schekman also encouraged his students to pursue outside interests, such as travel and the arts. It may not be surprising then, that Schekman served as the editor of *The Proceedings of the National Academy of Sciences* from 2006 until 2011, when he left to become editor-in-chief of the open-access journal *eLife*.

Ploegh says that the award also highlights the relevance of basic research to biomedicine. "This recognizes a fundamental discovery that I think was made without any urgent need to solve a medical problem."

All three laureates are previous winners of the Albert Lasker Basic Medical Research Award, which is often regarded as an indicator of future Nobel prizes in medicine. Südhof won it this year (along with Scheller) for his work on the release of neurotransmitters. Schekman and Rothman shared the award in 2002. ■

FABRICE COFFRINI/AFP/GETTY

Nobel laureates Peter Higgs (right) and François Englert at CERN in July 2012.

NOBEL PRIZE

Higgs theorists amass physics prize

François Englert and Peter Higgs rewarded with Nobel 50 years after hunt for boson began.

BY RICHARD VAN NOORDEN

Thousands of scientists were involved in hunting down the Higgs boson, this generation's greatest discovery in particle physics. But for the committee awarding the Nobel Prize in Physics, two names mattered most. In an announcement on 8 October in Stockholm, Peter Higgs of the University of Edinburgh, UK, and François Englert of the Free University of Brussels were named Nobel laureates for developing the theory of what is now commonly called the Higgs mechanism: the process by which a field pervading space gives other fundamental particles mass, and which implies the existence of the Higgs boson. Regarding the committee's choice, "I think in all honesty, this is what I would have done," says John Ellis, a theoretical physicist at CERN, Europe's particle-physics lab near Geneva, Switzerland.

The existence of the boson was announced to cheers at CERN on 4 July last year, after the particle was fleetingly produced in high-energy collisions at the lab's €3-billion (US$4.1-billion) Large Hadron Collider. It would have been too complicated to try to honour the experimenters with the Nobel, says Ellis, who joined other CERN theorists in popping open champagne as the award was announced. "Englert and Higgs' pioneering work richly deserved this prize," he adds.

"I'm very, very happy to have the recognition of this extraordinary reward," says Englert. Higgs, who is notoriously modest and suffered a bout of bronchitis last month, made himself unavailable for interviews. The two winners had met for the first time at CERN last July.

The Higgs boson was the missing piece in the standard model of particle physics, which describes all known fundamental particles and forces, apart from gravity. The boson itself is the smallest possible ripple of the Higgs field, which gives mass to particles including electrons, quarks and the W and Z bosons that carry the weak nuclear force.

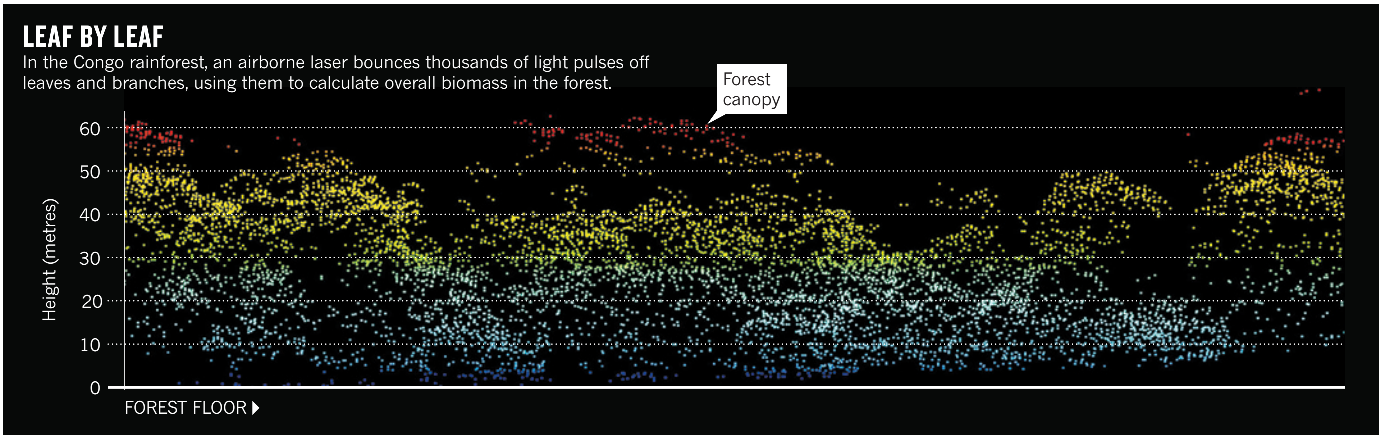
The idea was mooted in the 1960s, when physicists trying to describe the fundamental forces were wrestling with "embarrassing massless particles floating around in their theories", as Ellis puts it. In 1964, six physicists independently worked out how a field would resolve the problem. Robert Brout (who died in 2011) and Englert were the first to publish, in August 1964, followed three weeks later by Higgs — the only author, at the time, to allude to the heavy boson that the theory implied. Tom Kibble, Gerald Guralnik and Carl Hagen followed. "Almost nobody paid any attention," says Ellis — mostly because physicists were unsure how to make calculations using such theories. It was only after 1971, when Gerard 't Hooft sorted out the mathematics, that citations started shooting up and the quest for the Higgs began in earnest.

So numerous were the theorists involved, that Higgs reputedly referred to the ABEGHHK'tH (Anderson–Brout–Englert–Guralnik–Hagen–Higgs–Kibble–'t Hooft) mechanism. But that list of names is nothing compared with the legion of experimenters who joined the quest to track down the boson, with increasingly powerful particle accelerators that produced their own Nobel-prizewinning findings along the way.

"It's really an incredible thing that it's happened in my lifetime," Higgs told the audience at CERN when the particle was announced.

Alan Walker, a colleague of Higgs at Edinburgh, says, "That day was for the experimentalists. I guess today is for the theorists." ■

SOURCE: S. SAATCHI



DEFORESTATION

Congo carbon plan kicks off

Democratic Republic of the Congo maps forest biomass to attract carbon credits.

BY JEFF TOLLEFSON

After lifting off this week from the chaotic urban jungle of Kinshasa, scientists guided a twin-engine aircraft over the real jungle. With a small onboard laser, they began sweeping the vast rainforest of the Democratic Republic of the Congo (DRC), bouncing photons off leaves and branches. The aim is to measure — and perhaps to preserve — the carbon locked up in the tropical forests that cover two-thirds of the country.

The data will also enhance scientists' understanding of tropical forests' role in global climate regulation. "We know very little about the tropics," says Sassan Saatchi, a remote-sensing scientist at NASA's Jet Propulsion Laboratory in Pasadena, California, who is leading the data analysis for the project. "If these countries know how to monitor their forests quantitatively, it will help us solve the problem."

The flights are part of a two-year, €6-million (US$8-million) project, funded by Germany, to combat the kind of tropical deforestation that today accounts for up to 15% of the world's carbon dioxide emissions. For years, governments have debated a treaty that would allow developed countries to offset their own greenhouse-gas emissions by helping tropical countries to conserve their forests. Negotiations have stalled, but wealthy nations have already pledged several billion dollars to get the scheme started.

The DRC is one of several tropical countries hoping to attract some of that money. But first, it must map biomass across a staggering 155 million hectares — the largest tropical forest in a single country outside Brazil — and then establish a system to monitor deforestation from space. "The country is so huge, it has so much forest, and it's so poor," says Aurélie Shapiro, the project's technical lead with WWF Germany, an environmental group that is managing the project with the DRC. "If the Democratic Republic of the Congo can do it, anybody can."

To help, WWF Germany turned to Saatchi, who has already mapped out carbon across the world's tropics, but at relatively low resolution of 1 kilometre. His calculations suggest that the DRC's forests contain 22 billion to 24 billion tonnes of carbon, equivalent to more than double the greenhouse gases emitted last year. But the measurements need to be more precise, he says, because biomass can vary by up to 50% between patches of forest. With a map at 100-metre resolution, the DRC will be better able to calculate the emissions that would result from clearing a block of forest. That, in turn, would be key information for companies or governments seeking to offset their emissions by protecting DRC land.

> ***"We have only limited areas where it's safe to land."***

The project will start with the laser data collected by planes flying out of Kinshasa (see 'Leaf by leaf'). The team will then begin operating out of more remote airports, many little more than dirt runways in the jungle. The plan is to sample at more than 200 locations, selected to capture different types of forest — and to avoid conflict zones in the war-torn country. It is a daunting task that has required more than two years of planning. "We have to carry our fuel with us," says Saatchi. "And we have only limited areas where it's safe to land."

The aerial measurements will cover just half a million hectares, but can be extrapolated to the entire forest through calibration with ground data and satellite imagery. The data are first cross-checked with measurements of tree girth and height at several hundred ground plots — the classic way to calculate forest biomass. The biomass estimates are then used to calibrate imagery from NASA's Landsat spacecraft and radar data from Japan's Advanced Land Observing Satellite, to produce a measure of carbon for all 155 million hectares of jungle.

The DRC is not the only country embarking on large-scale forest measurements. Gabon, in the western Congo Basin, is planning to map its forests using the same technique, and in the past few years Peru has mapped the carbon in its part of the Amazon (see *Nature* **461,** 1048–1052; 2009). Brazil has made less-detailed assessments of forest carbon, but its system for monitoring deforestation is the world's most advanced. Scientists with the country's National Institute for Space Research in São José dos Campos are now helping the DRC to set up a similar system, based on freely available Landsat data, to track deforestation — and ultimately to verify reductions in such losses and sell carbon offsets.

Matthew Hansen, a remote-sensing scientist at the University of Maryland in College Park who works with forest mappers in Kinshasa, says that the DRC project faces hurdles. Much of the deforestation — by urban dwellers seeking wood and charcoal, and by farmers clearing small plots of land — is small-scale and difficult to see by satellite. Constant clouds also complicate the kind of monitoring that Brazil has pioneered.

Large-scale deforestation, which would be easier to monitor, has yet to take off in the Congo Basin, owing in part to political instability. But Hansen fears that could change if investments increase in agriculture projects such as oil-palm plantations. "If that ever happens, you could see some dramatic changes," he says. ■

PATRICK SEMANSKY/AP

The National Security Agency is the largest employer of mathematicians in the United States.

SECURITY

Researchers split over NSA hacking

Cryptographers condemn US National Security Agency's tapping and tampering, but mathematicians shrug.

BY ANN FINKBEINER

The US National Security Agency (NSA) has upset a great many people this year. Since June, newspapers have been using documents leaked by former intelligence worker Edward Snowden to show how the secretive but powerful agency has spied on the communications of US citizens and foreign governments. Last month, the media reported that the NSA, which is based in Fort Meade, Maryland, had undermined Internet security standards. The revelations have sparked international outrage at the highest levels — even the president of Brazil cancelled a visit to the United States because of the spying.

Yet amid the uproar, NSA-supported mathematicians and computer scientists have remained mostly quiet, to the growing frustration of others in similar fields. "Most have never met a funding source they do not like," says Phillip Rogaway, a computer scientist at the University of California, Davis, who has sworn not to accept NSA funding and is critical of other researchers' silence. "And most of us have little sense of social responsibility."

Mathematicians and the NSA are certainly interdependent. The agency declares that it is the United States' largest maths employer, and Samuel Rankin, director of the Washington DC office of the American Mathematical Society, estimates that the agency hires 30–40 mathematicians every year. The NSA routinely holds job fairs on university campuses, and academic researchers can work at the agency on sabbaticals. In 2013, the agency's mathematical sciences programme offered more than US$3.3 million in research grants.

Furthermore, the NSA has designated more than 150 colleges and universities as centres of excellence, which qualifies students and faculty members for extra support. It can also fund research indirectly through other agencies, and so the total amount of support may be much higher. A leaked budget document says that the NSA spends more than $400 million a year on research and technology — although only a fraction of this money might go to research outside the agency itself.

Many US researchers, especially those towards the basic-research end of the spectrum, are comfortable with the NSA's need for their expertise. Christopher Monroe, a physicist at the University of Maryland in College Park, is among them. He previously had an NSA grant for basic research on controlling cold atoms, which can form the basis of the qubits of information in quantum computers. He notes that he is free to publish in the open literature, and he has no problems with the NSA research facilities in physical sciences, telecommunications and languages that sit on his campus. Monroe is sympathetic to the NSA's need to track the development of quantum computers that could one day be used to crack codes beyond the ability of conventional machines. "I understand what's in the newspapers," he says, "but the NSA is funding serious long-term fundamental research and I'm happy they're doing it."

Dena Tsamitis, director of education, outreach and training at Carnegie Mellon University's cybersecurity research centre in Pittsburgh, Pennsylvania, also wants to maintain the relationship. She oversees visitors and recruiters from the NSA but her centre gets no direct funding. She says that her graduate students understand the NSA's public surveillance to be "a policy decision, not a technology decision. Our students are most interested in the technology." And the NSA, she says — echoing many other researchers — "has very interesting technology problems".

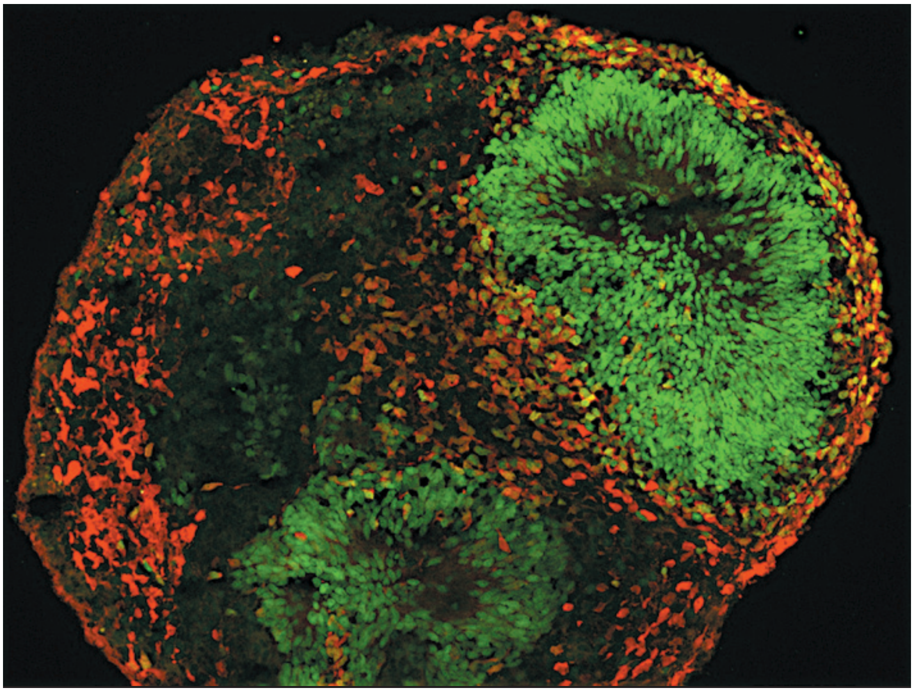
The academics who are professionally uneasy with the NSA tend to lie on the applied end of the spectrum: they work on computer security and cryptography rather than pure mathematics and basic physics. Matthew Green, a cryptographer at Johns Hopkins University in Baltimore, Maryland, says that these researchers are unsettled in part because they are dependent on protocols developed by the US National Institute of Standards and Technology (NIST) to govern most encrypted web traffic. When it was revealed that the NSA had inserted a 'back door' into the NIST standards to allow snooping, some of them felt betrayed.

"I understand what's in the newspapers, but the NSA is funding serious long-term fundamental research and I'm happy they're doing it."

"We certainly had no idea that they were tampering with products or standards," says Green. He is one of 47 technologists who on 4 October sent a letter to the director of a group created last month by US President Barack Obama to review NSA practices, protesting because the group does not include any independent technologists.

Edward Felten, who studies computer security at Princeton University in New Jersey, says that the NSA's breach of security standards means that cryptographers will need to change what they call their threat model — the set of assumptions about possible attacks to guard against. Now the attacks might come from the home team. "There was a sense of certain lines that NSA wouldn't cross," says Felten, "and now we're not so sure about that." ■

SOURCE: ANCA PASCA



Pluripotent stem cells generated from patients' skin cells could be used to personalize brain therapies.

NEUROSCIENCE

Novartis reboots brain division

After years in the doldrums, research into neurological disorders is about to undergo a major change of direction.

BY ALISON ABBOTT

In a sign that psychiatric-disease research is entering a new era, the pharmaceutical giant Novartis has hired an expert in neural circuitry, rather than pharmacology, to head its relaunched neuroscience division.

The appointment of 42-year-old Ricardo Dolmetsch, who has spent his entire career in academic research, signifies a radical policy shift for the company, as it moves away from conventional neurotransmitter research to concentrate on analysing the neural circuitry that causes brain diseases. The decision suggests Novartis is confident that after years of fruitless research in the field, revolutionary advancements in, for example, genetic and stem-cell technologies will pay dividends. The company intends to hire 100 new staff members for the department over the next 3 years. But the move is risky: even if it pans out, new drugs for common disorders such as schizophrenia could be decades away from reaching the market.

Dolmetsch, a former senior director at the Allen Institute for brain Science in Seattle, Washington, who has also worked at Stanford University School of Medicine in California, says that his new role gives him access to previously unimaginable resources. "I had this idea that big pharma was a slow, plodding, conservative giant," he says. "I was surprised by the depth of science at Novartis."

An expert in autism spectrum disorder, he was also attracted by the prospect of contributing to the development of therapies — something that academic institutions are poorly equipped to do — particularly because one of his own sons has autism. There was "not much enthusiasm" for studying disease at the Allen Institute, which focuses instead on basic research into brain science, he says.

The World Health Organization estimates that one-quarter of people develop one or more disabling mental illnesses during their lifetime. Yet in the past few years, most of the world's major drug companies — Novartis included — have abandoned psychiatric disease.

The company enjoyed some successes in the field in the 1960s and 1970s, including the development of clozapine, the first

of the 'atypical antipsychotic' drugs, and imipramine, the first of the tricyclic class of antidepressants. Both of these drugs spawned many look-alikes. But since these early breakthroughs, no radical improvements in therapy have emerged. For example, there is still no drug for the core symptoms of autism or for the particular cognitive deficits in schizophrenia that prevent most sufferers from holding down a job.

When Novartis closed down its conventional drug-discovery programme in Basel, Switzerland, last year (see *Nature* **480,** 161–162; 2011), it was already planning to reopen with an entirely new approach at its global drug-discovery headquarters, the Novartis Institutes for BioMedical Research (NIBR) in Cambridge, Massachusetts. In its sights are neurodevelopmental diseases such as autism, schizophrenia and bipolar disorder, as well as neurodegenerative diseases such as Parkinson's and Alzheimer's disease.

Most current psychiatric drugs are based on the modulation of neurotransmitters, but the company's plan is instead to focus on connections between neurons, with the aim of addressing the precise neural circuitry that goes wrong in psychiatric disease, says Mark Fishman, the president of the NIBR. Genetics and environment both strongly influence how disease-causing circuits establish themselves.

Increasingly powerful genetic studies have begun to identify genes that might cause or contribute to the disorders (Cross-Disorder Group of the Psychiatric Genomics Consortium *Nature Genet.* **45,** 984–994; 2013), and scientists can now efficiently model many of these defects in animals. With optogenetics — a technique introduced in 2005 that uses genetics in concert with light to activate or inhibit particular neurons — they can identify in living animals the neuronal circuits that the genes affect, and thus which genes could be targeted for therapy. In addition, stem-cell technologies now allow scientists to turn skin or hair-follicle cells from individual patients back into pluripotent stem cells, differentiate them into neurons and then study the connections that the neurons make with one another. Such induced pluripotent stem cells might one day aid the personalization of therapies to individual patients.

"We'll begin to see a lot of clinical trials within a decade or so."

All of these technologies are young. Novartis is planning to collaborate closely with academic institutions to generate more genetic clues, build disease models and deepen the understanding of neural circuits in health and disease. The science will be published, says Dolmetsch, but Novartis will have first access to mutations and circuitry that are found to be linked to a disorder, for use in the company's drug-development programme. Novartis will also have the freedom to develop drugs around any target that emerges from genetic screens that it co-sponsors. "If academic groups feel they should patent targets, and if we contributed to the screen that revealed them, then we want the option to license those patents," says Dolmetsch.

Finding good targets, and moving from a target to a safe and effective drug, will not happen overnight, Dolmetsch stresses. "Everyone realizes that this is going to take a long time to bear fruit," he says. Steve Hyman, director of the Stanley Center for Psychiatric Research in Cambridge, Massachusetts, guesses that in five years — "if we all work hard enough" — it will be clear whether or not the new approaches are likely to be useful.

Hyman, who is an adviser to Novartis and a former director of the US National Institute of Mental Health, predicts that for diseases caused by a single gene, such as some rare but devastating forms of autism or intellectual disabilities, "we'll begin to see a lot of clinical trials within a decade or so".

The vastness of the task is keeping most big drug companies out of the game, although Roche, based in Basel, is committed to a similar approach, says the company's spokesman Stepan Kracala.

The European Commission is also trying to foster a collaborative approach through its Innovative Medicines Initiative. Two of its multimillion-euro programmes — EU-AIMS (European Autism Interventions — a Multicentre Study for Developing New Medications) and NEWMEDS (Novel Methods leading to New Medications in Depression and Schizophrenia) — combine different pharmaceutical companies and academic groups in projects that probe genetic and neural circuitry in autism, schizophrenia and depression.

Dolmetsch doesn't feel that his move to Novartis has distanced him from the intellectual centre of things. "The science at the NIBR is arguably better and certainly more rigorous than in most of academia," he says. ■ SEE EDITORIAL P.141

MEDICAL RESEARCH

Secrets of trial data revealed

Records of side effects seen in clinical tests are missing from publicly available documents.

BY DANIEL CRESSEY

Drug companies, regulatory bodies and research groups are in a deadlock over plans to throw open the vaults of sensitive data gleaned from clinical trials. But just as the battle seems to be coming to a head, a study (B. Wieseler *et al. PLoS Medicine* **10,** e1001526; 2013) has revealed exactly why the restricted information could be so valuable to researchers.

According to the analysis, crucial trial information, such as mortality rates and serious side effects, is missing from much published data. But it can frequently be found in standard non-public documents prepared by industry, known as clinical study reports (CSRs). Missing information uncovered by the study includes details of depression symptoms in trials for antidepressant drugs, and details of heart attacks and strokes in diabetes-drug trials. "These are very, very important variables and outcomes," says the study's lead author, Beate Wieseler, head of the drug-assessment department at the Institute for Quality and Efficiency in Health Care in Cologne, Germany.

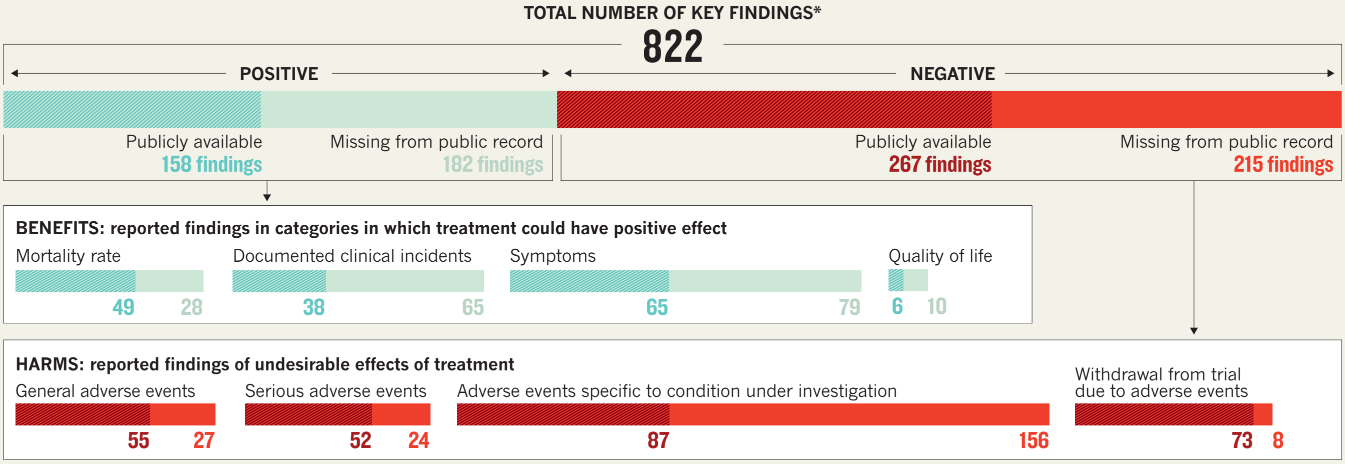
Wieseler strongly supports making all CSRs publicly available. "That is not an option, but a necessity," she says. "There should be no question that these documents be made available."

The issue of CSR access is shaping up to be a key sticking point in a major European push to make available more data from clinical trials. Drug companies send the reports to the European Union's European Medicines Agency (EMA) when they apply for licences to sell their products, and the agency has made clear that it would like to release the forms as part of a drive to increase transparency. It is currently drawing up a policy to that effect. Advocates of transparency say that such a policy will allow greater scrutiny of the benefits and harms of drugs, and will help researchers who are working on cures for diseases.

Some companies, including pharmaceuticals giants Roche, based in Basel, Switzerland, and GlaxoSmithKline in London, have already said that they will make their CSRs available to vetted researchers. But industry has pushed back against the wider moves to greater transparency.

INSIDE DRUG-COMPANY DATA VAULTS

In data from 86 clinical trials, investigators recorded 822 important findings, ranging from effect on mortality rate to fatal heart attacks. A review compared the information that was made publicly available with that available only in unpublished drug-company reports. The public information was found to omit crucial facts.



*For trials with both unpublished clinical study report (CSR) and public record; only findings with complete information shown.

SOURCE: B. WIESELER *ET AL. PLOS MEDICINE* **10**, E1001526; 2013

In an e-mail to *Nature*, Richard Bergström, the director-general of the European Federation of Pharmaceutical Industries and Associations (EFPIA), said that in their current state, CSRs are fundamentally unsuitable for publication. The EFPIA believes that the EMA's proposed mechanisms for protecting the commercially sensitive and personal details contained in many of these documents are inadequate — the mechanisms require that information be 'de-identified' and released only for bona fide research purposes. Such personal data must be properly redacted if these documents are to be made public, says Bergström.

"My members are very concerned about this," he says. "If the EMA accepts our redactions, we have no problem." If the agency disregards the EFPIA's concerns, however, Bergström warns that there may be a series of lawsuits against the EMA.

The EMA has already been taken to court by two biotechnology companies, AbbVie of North Chicago, Illinois, and InterMune of Brisbane, California, to block release of their information under existing rules, whereby any researcher can request information. The cases are ongoing and have severely restricted releases of other requested data.

But advocates of greater transparency are fighting for the EMA to release all data, which they say can easily be anonymized.

In the study, Wieseler and her team examined the information in 101 CSRs given to them by pharmaceutical companies, and compared it with the information about the same clinical trials that is available in the public domain, found in journal publications and reports in trial registries.

The unpublished sources provided considerably more information on the key findings, or outcomes, relating to the treatments being assessed, including mortality rates and adverse events. In total, the team found 1,080 outcomes relevant to patients. Complete information was available for 86% of these in the CSRs, but for only 39% in the publicly available information. When looking only at outcomes related to harm — such as adverse reactions to a drug — the researchers found that complete information was found for 87% of outcomes in the CSRs, but for only 43% in public records. In some cases, no public information was available at all for a given CSR. But when looking only at cases in which there was both a CSR and a public document, the researchers found similar proportions of missing information in the publicly available records (see 'Inside drug-company data vaults').

"What's portrayed in the journal world does not reflect the whole truth."

"Does it surprise me? No. Is this stuff really important? Yes," says Carl Heneghan, director of the Centre for Evidence-Based Medicine at the University of Oxford, UK, and co-founder of a group called AllTrials, which campaigns for more clinical-trial transparency. "It's becoming obvious that what's portrayed in the journal world does not reflect the whole truth."

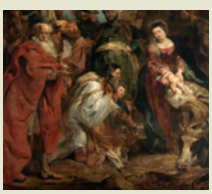
In a public consultation that ended last week, the EMA's proposals attracted support from other quarters. The Medicines and Healthcare Products Regulatory Agency, the UK drug regulator, says that it welcomes the EMA's proposals, "and will work to align our position with theirs".

A joint response from UK biomedical-research funders, including the Wellcome Trust and the Medical Research Council also backs the EMA's plans. However, the group voices concerns over the sharing of data from individual patients; it says that more safeguards should be put in place to ensure that data are given only to trusted researchers who will not "wrongfully contradict" the results of trials.

The EMA hopes to have its policy in place by the start of next year. Documents without sensitive personal data will be made available for download from its website, whereas those that do contain such data will be available only after anonymization to vetted researchers. Documents with commercially confidential information will be released separately. ■

MORE ONLINE

PICTURE STORY

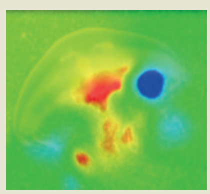


Paintings turning black? Blame mercury go.nature.com/upcbhe

TOP STORIES

- 'Higgsogenesis' proposed to explain dark matter go.nature.com/rbeihq
- Formula forecasts the probable number of citations for research papers go.nature.com/pnz7db
- Amorous insects predict the weather go.nature.com/mq3a28

VIDEO OF THE WEEK



Why a jellyfish is the ocean's most efficient swimmer go.nature.com/kyqdov

PETER PAUL RUBENS, *ADORATION OF THE MAGI*, KMSKA, LUKAS-ART IN FLANDERS, HUGO MAERTENS

BRAD GEMMELL

ILLUSTRATION BY ADAM NICKEL

THROUGH THE EYES OF A MOUSE

Some brain researchers are increasingly using mice to study visual processing, but others fear the move is short-sighted.

BY MONYA BAKER

When Cris Niell said that he wanted to study how mice see, it did not go over well with more-senior neuroscientists. Mice are nocturnal and navigate largely using their noses and whiskers, so many researchers believed that the nursery rhyme — *Three Blind Mice* — was true enough to make many vision experiments pointless. The obvious alternative model was monkeys, which have large, forward-looking eyes and keen vision. What's more, scientists could rely on decades of established techniques using primates, and it is relatively straightforward to apply the results to the human visual system. "People were saying, 'studying vision in mice, that's crazy,'" Niell recalls.

But he was convinced that the rodents offered unique opportunities. Since the 1960s, researchers have used cats and monkeys to uncover important clues about how the brain turns information from the eyes into images recognized by the mind. But to investigate that process at the cellular level, researchers must be able to manipulate and monitor neurons precisely — difficult in cats and monkeys, much easier in mice. If mice and primates turned out to process visual stimuli similarly, Niell thought, that discovery could unleash a torrent of data about how information is extracted from stimuli — and even, more generally, about how the brain works.

He found a rare supporter in Michael Stryker at the University of California, San Francisco, who had already seen his share of crazy experiments in mouse vision. Stryker offered Niell a postdoctoral position in his lab, and the pair began setting up experiments in 2005.

Nearly a decade later, the two researchers are in better company. At last year's annual meeting of the Society for Neuroscience, Niell attended packed sessions on mouse vision. In March 2012, the Allen Institute for Brain Science in Seattle, Washington, announced a ten-year plan to spend more than US$100 million to map mouse visual areas. And in June this year, the curriculum of a two-week course on vision at Cold Spring Harbor Laboratory in New York featured mice front and centre. More than three-quarters of the 22 students investigating how the visual system works were using mice, says course co-director Andrew Huberman, a neuroscientist at the University of California, San Diego, who has worked on animals from cuttlefish to macaques. In 2001, he says, there may have been a student or two using mice to study how the visual system develops, but no one was studying function. "It's an explosion."

The surge of interest stems largely from advances that give researchers the ability to monitor and control specific mouse neurons using light. Logistical and ethical considerations are also a big draw. Studies with mice are much cheaper, faster and less likely to raise moral concerns than work with monkeys.

But whether they will reveal useful information about human vision is very much an open question, says Huberman. "The mouse visual cortex is like the smartphone of neuroscience," he says. "Everyone feels the need to get one to play with, but it still remains to be seen if it's merely a convenience, a colossal distraction or the greatest thing since the discovery of electricity."

"THE MOUSE VISUAL CORTEX IS LIKE THE SMARTPHONE OF NEUROSCIENCE. EVERYONE FEELS THE NEED TO GET ONE, BUT IT REMAINS TO BE SEEN IF IT'S A CONVENIENCE OR A DISTRACTION."

DRAWN TO RODENTS

Niell wanted to revisit some of the most well-known and seminal experiments in vision science. In the 1950s and 1960s, David Hubel (who died late last month) and Torsten Wiesel pushed electrodes into the backs of cat and monkey brains, and patched the signals to a speaker to track the activity of neurons. The researchers listened in as the animals viewed tilted lines and moving dots; the crackles they heard revealed that organized regions of neurons respond to motion and edges[1]. The results, which later earned Hubel and Wiesel a Nobel prize, became a canonical example of 'cortical computation', in which interconnected neurons transmit and transform information.

It turns out that neurons in the visual cortex process input from the eyes extremely selectively: some respond only to vertical lines, others to horizontal ones, still others to stripes that tilt 40° to the left, to dots creeping up 30° to the right, and so on.

Niell, who now runs a lab at the University of Oregon in Eugene, knew that it would not be easy to discover whether these findings would hold true for mice. The electrodes available at the time often damaged neurons in fragile mouse brains, disrupting activity rather than monitoring it. But after revising their procedures and redesigning their equipment, Niell and Stryker worked out a way to record from individual mouse brain cells using silicon microprobes.

Stryker, who had trained with Hubel and Wiesel, recalls seeing the first graph that plotted how a neuron responded to a mouse viewing a series of tilting lines. The graph showed sharp, narrow peaks of neural activity at specific orientations[2]. If the lines were tilted just 20° from the preferred angle, the cell fell silent. "I just couldn't believe how pretty it was," says Stryker. "It was like the figure in a book."

The experiments showed that neurons in the mouse visual cortex are about as selective as those in the cat or monkey brain. Niell and Stryker considered that to be strong evidence that the mouse could be used as a model for visual processing in higher animals.

When word got out, the team soon had visitors. Among the earliest were Hillel Adesnik and Bassam Atallah, two neuroscientists then working in the lab of Massimo Scanziani at the University of California, San Diego. Adesnik and Atallah had been studying dissected slices of mouse brain to catalogue how subtypes of neurons were connected. They could test how cells in the freshly dissected tissue responded when stimulated with electricity, but it was hard to know what kind of processing, if any, happens in a brain slice. They really wanted to probe how brain circuits responded to real physiological stimuli — the type delivered by the eyes, ears, nose or skin of a living mouse. When they heard of the results from San Francisco, they hopped on their motorcycles and rode 800 kilometres to learn Niell and Stryker's technique.

➲ NATURE.COM
For more on neuroscience in mice, see:
go.nature.com/ztofam

Since that visit, Adesnik, Atallah and others have done experiments that show how neurons interact in intact circuits in the mouse visual cortex. Such work is beginning to reveal how subtypes of neurons cooperate to extract information about the world; it also hints at which stimuli mice notice or ignore. "I consider the mouse visual cortex as the first pillar in a bridge that will link cellular and systems neuroscience," says Scanziani.

It is too early to know how far that bridge will extend, says Edward Callaway, a neurobiologist at the Salk Institute for Biological Sciences in La Jolla, California. "So far, we haven't learned anything fundamentally new from the mouse visual system. And that's not surprising, since we've studied monkeys for the past 40 years."

But the early data are fuelling mouse researchers' hopes that this simple, easily manipulated model could shed light on more-complex brains. For instance, in a sign that mice have sophisticated processing, researchers showed that, like primates, mice have visual areas that receive input beyond the primary visual cortex[3].

Around the same time that Niell and Stryker were learning that mice can see stripes, Matteo Carandini at University College London switched to mice after years of researching vision in monkeys and cats. He wanted to study neural circuitry in the context of behaviour, but for that he had to record from inside individual neurons. That is hard to do in monkeys — and although it can be done in cats, they baulk at the behavioural tasks required. So Carandini began working out training programmes to let him explore what patterns mice perceive as they move around, and how they act on their perceptions.

He and his team developed one task in which mice press a button when they see stripes. The team also monitored mouse visual processing as the animals ran on a treadmill or explored a virtual environment. Carandini now wants to manipulate particular neurons during such experiments to see how the mouse's behaviour changes.

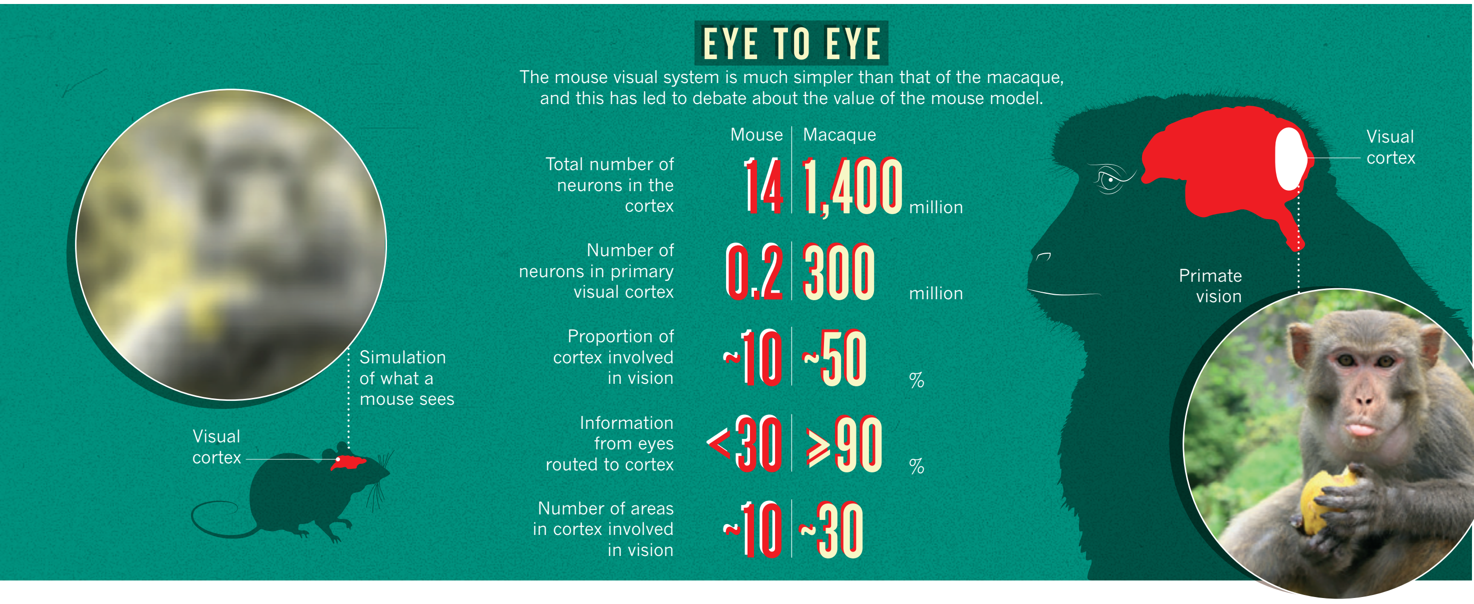
He wants to learn how parts of the brain cooperate: he thinks that work on the mouse's small, flat brain could shed light on processes in higher animals, such as perception and decision-making, and how these are affected by distraction (such as noise) or motivation (such as thirst). "The frontier at this moment is understanding how different parts of the cortex work together," he says. "This is a problem that any animal with a cortex has."

A BLURRY PICTURE

Carandini and others who work on mouse vision all acknowledge that there are limits to this line of research. No one denies that mice see poorly; Niell estimates that they have the equivalent of 20/2,000 human vision (which would qualify them as legally blind). The general rule of thumb is that mouse eyesight is about as good as what humans see in their far-off peripheral vision.

So there will be tasks that rodents will not be able to perform — particularly those related to aspects of facial recognition and visual attention. "To really get to behaviour in a more meaningful way, we'd probably have to use primates," says Callaway, who is working with mice as well as improving genetic tools for studying monkeys.

Still, enthusiasts say, the similarities between mice and humans



PHOTOGRAPH: LASZLO ILYES/FLICKR
MOUSE VISION SIMULATION: JOHANNES BURGE/WILSON GEISLER/UT AUSTIN

outweigh the differences. The mouse visual cortex contains the same neural subtypes as the human visual cortex, in about the same proportions, and the subtypes seem to hook up to each other using the same rules. Evolutionarily, mice are more closely related to humans than are cats. And the fact that a mouse has fewer brain regions than either a primate or a cat, and can distinguish a smaller set of possible images, makes it more experimentally tractable, says Carandini. "The basic rules of computation, I believe, are more general and canonical. Our chances of discovering them are much better in a mouse."

Not all neuroscientists embrace the mouse as a model. "I don't think there's good evidence that mice use the visual cortex the way primates do. Or that the mouse visual cortex is organized the way that primates' are, or that mice are a good model for vision," says Tony Movshon, a neuroscientist at New York University.

The most obvious difference is size: the entire mouse brain is one-fiftieth as big as the part of the macaque brain devoted to vision alone. A macaque's primary visual cortex has more than 1,000 times as many neurons as a mouse's, and a much greater fraction of the macaque brain is devoted to vision (see 'Eye to eye'). Primates have plenty of neurons in dozens of visual areas with targeted purposes, such as recognizing faces and tracking motion. By contrast, the visual regions identified in mouse brains are "tiny little patches of cortex", says Movshon; they extend for micrometres and millimetres, rather than centimetres. The areas are simply too small for the extensive, regional communication observed in primate visual areas, he says. "They can't work the same way."

BUSY BRAINS

Perhaps the biggest problem is that the mouse visual cortex performs many functions besides vision, so the systems that support visual processing could be fundamentally different from those in the primate brain. It is like studying heart function in some alien organ that not only pumps blood but also takes on the respective gas-exchange and electrolyte-balancing functions of the lungs and kidneys.

For all these reasons, instead of switching his vision studies from monkeys to the more tractable mice, Movshon is putting his faith in nascent efforts to take the tools that work so well in mice and adapt them to manipulate subtypes of neurons in monkey brains. "What people are doing now is to pretend that the mouse is a tiny monkey with a pointy noise and whiskers and to hope for the best," he says. Paul Martin, a vision scientist at the University of Sydney in Australia, agrees that scientists could encounter severe problems when they try to relate mouse data to the human experience. "A shopping trolley and a Formula 1 motor car both have wheels and obey Newton's laws of motion. But the motor — what makes them move — is quite different," he says.

Nicholas Priebe at the University of Texas at Austin is advocating more comparative studies to tease out how processing differs between mouse brains and those of other species — and how it is similar. This year, he reported striking differences in how brain regions in cats and mice contribute to selectivity[4]. Discrepancies between mice and primates do not mean that mouse brains have nothing to reveal about human brains, he says, but scientists need to proceed with as much caution as enthusiasm. "If you try to apply everything you learn from the mouse to our brain, then I think there's a serious problem," says Priebe.

For most, the debate is not about whether to study visual processing in the mouse cortex, but about what questions will also apply to higher animals. Many researchers are hopeful that mouse experiments will generate hypotheses for primate research, and that comparable experiments could one day go back and forth between animal systems. But after decades of focusing on primates, researchers have some catching up to do in rodents, says Callaway. "We can't begin to do those things in a mouse until we know more about what they use the visual system to do."

That is part of the aim of the initiative by the Allen Institute for Brain Science to map the mouse visual cortex and visual processing areas. The goal, says Clay Reid, who co-leads the project, is to start at the bottom, building up to big questions about how the brain works. Reid and his team plan to catalogue the cell types and connections in the mouse's visual area, and to monitor what happens as the animals look at, and respond to, a stimulus. Then the team will see how responses change when particular neurons are suppressed or activated. Such experiments are about more than just vision. "We are doing it to try to understand principles of cortical computation and the relationship between cortical activity and behaviour," says Reid. Armed with some clues to those processes, scientists will be able to test these hypotheses in other animals.

Eight years after his initial experiments, Niell is glad to see more researchers embracing what the mouse has to offer. Of course, such a simple system cannot answer every question we have about the human brain, but researchers should learn what they can, says Niell. "You can make so much headway with a mouse that it's silly not to." ■

Monya Baker *is a reporter and editor with* Nature *in San Francisco, California.*

1. Hubel, D. H. & Wiesel, T. N. *J. Physiol. Lond.* **160,** 106–154 (1962).
2. Niell, C. M. & Stryker, M. P. *J. Neurosci.* **28,** 7520–7536 (2008).
3. Wang, Q. & Burkhalter, A. *J. Comp. Neurol.* **502,** 339–357 (2007).
4. Scholl, B., Tan, A. Y. Y., Corey, J. & Priebe, N. J. *J. Neurosci.* **33,** 10616–10624 (2013).

NEWS FEATURE

Footprints in the *forest*

Researchers are tracking just how much impact ancient peoples had on the Amazon.

BY JEFF TOLLEFSON

Indigenous populations may have long maintained plantations of palms and bananas.

Crystal McMichael first led a crew into the Amazon jungle in 2007, looking for signs of ancient human disturbances. Armed with machetes, they hacked their way through thick vegetation while fending off spiders, mosquitoes and bees. They were exploring around Ecuador's Lake Ayauchi, which McMichael knew held the earliest record of maize (corn) cultivation in the Amazon from around 6,000 years ago. But the jungle hid its secrets well. "If you looked at the forest you wouldn't realize there was any ancient disturbance," says McMichael, now a research scientist at the Florida Institute of Technology in Melbourne. "You have to dig."

Scientists have struggled for decades to uncover humanity's historical footprint in the forest and determine what kind of an impact people had centuries to millennia ago. Their goal is to understand the evolution of the rainforest and just how much of the landscape we see today is 'natural' versus how much has been shaped by human hands.

Studies dating back to the 1950s suggested that small indigenous tribes merely scratched out a living in primitive villages before the arrival of Europeans. But more recently, researchers have proposed that the Amazon hosted complex societies that turned swathes of the forest into farms and orchards. Some estimates place the prehistoric population of the Amazon as high as 10 million — a huge number considering that the current population is around 30 million. The debate is heated. When McMichael and her colleagues reported[1] last year that indigenous occupations might have been rare in the most-remote parts of the jungle, their paper outraged archaeologists.

The topic evokes strong emotions in part because it touches on the sensitive issue of indigenous land claims and goes to the heart of conservation philosophy. If prehistoric human populations were limited and today's Amazon is relatively pristine, then one might assume that this otherwise stable and natural ecosystem would be altered by any human disturbance — let alone the clearance of vast tracts of forest for agriculture (in Brazil alone, an area greater than Germany has been cleared over the past 25 years). By contrast, if the primeval Amazon was filled with people who managed the landscape, then the forest might be capable of absorbing further human impacts. Encouraging indigenous practices, even on a large scale, might allow people to live in balance with the rainforest.

"The people who refuse to accept the human role are never going to understand how the environment that we appreciate today came to be," says anthropologist Clark Erickson of the University of Pennsylvania in Philadelphia, who believes that people were widespread throughout the Amazon. "And if you don't understand that, you will never know how to manage it."

HOSTILE LAND

When researchers started studying the Amazon, they entered what had long been viewed as an impenetrable and hostile environment. Scientists such as the late Betty Meggers, an archaeologist at the Smithsonian Institution in Washington DC, argued in the 1950s that the region's lush flora grew atop a layer of poor

RICKEY ROGERS/REUTERS/CORBIS

Networks of roads connect villages in the virgin forest.

TON KOENE/VISUALS UNLIMITED/CORBIS

soil that was unfit for cultivation and therefore large-scale civilization[2]. This theory fitted neatly within an old colonial paradigm that portrayed the Amazon as a largely empty jungle open to occupation and exploitation. Meggers documented pottery shards, burial sites and a network of defensive mounds on the island of Marajó at the mouth of the Amazon River. But, she argued, this community was short-lived owing to environmental limitations such as poor soil. This probably prevented large-scale development throughout the basin, she wrote: "Its levelling effect appears to be inescapable."

This view came under attack in the 1980s, first by Anna Roosevelt, an archaeologist at the University of Illinois at Chicago. On the island of Marajó, her work revealed a culture that had endured for nearly 1,000 years until about 400 AD — more than long enough to put theories about environmental limitations in doubt[3]. By the time Roosevelt published a detailed book[4] about her work on Marajó in 1991, the tide of opinion had begun to turn.

While Roosevelt was studying Marajó, William Balée, an anthropologist at Tulane University in New Orleans, Louisiana, was spending time with the Ka'apor tribe in the southeastern Amazon. In 1993, he documented the group's knowledge and use of a forest region that contained an unusually high concentration of useful species[5]. Tribal members referred to neighbouring patches of forest, by contrast, as wild and undisturbed. For Balée, this was a sign that many parts of the landscape were at one point cultivated as complex orchards. "Part of this was a backlash against the idea that natives tiptoed through the forest and left no footprints," says Erickson.

Since then, researchers working across the eastern and central Amazon have found deposits of '*terra preta*' (literally 'black earth' in Portuguese), which are fertile soils that are thought to have been created through cycles of fire and cultivation. Further earthworks, including mysterious systems of ditches and mounds, were uncovered throughout the 1990s in the western Amazon. By the mid-2000s, researchers had come to believe that prehistoric people once occupied large areas; built networks of roads, canals and bridges; cultivated crops such as maize and manioc (cassava); and maintained plantations of useful trees such as bananas and palms[6]. "These societies were fully on par with small to medium-sized populations anywhere on the globe in the 1400s," says anthropologist Michael Heckenberger at the University of Florida, Gainesville. "They weren't backward in any way."

> **NATURE.COM**
> View a slide show of more images online.
> go.nature.com/sbekpn

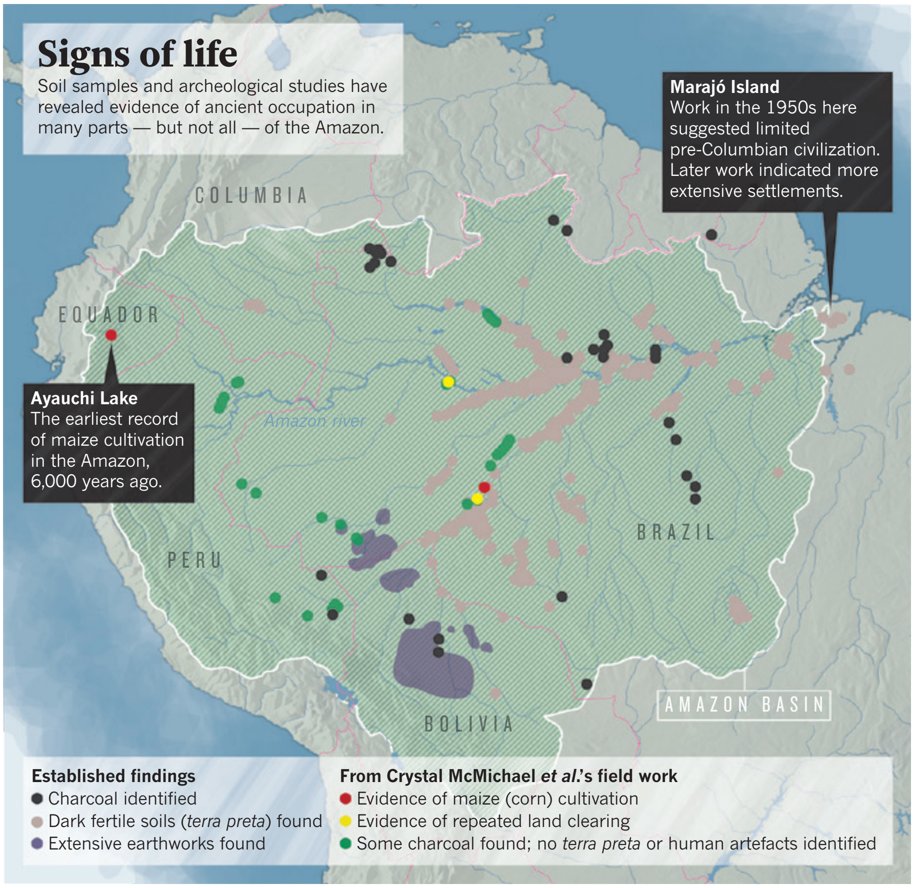
After such a radical shift in thinking, it was perhaps inevitable that the scientific pendulum would swing back the other way. In June 2012, a team of researchers led by McMichael and Mark Bush at the Florida Institute of Technology published a paper[1] arguing that human civilization was sparse across the wetter forests of western and central Amazonia. The team had collected 247 soil cores from dozens of sites and found charcoal in many locations — a sign of human fires. But none of the sites held human artefacts or *terra preta* (see 'Signs of life'). The team documented maize cultivation in only one instance, and just a couple of other sites revealed signs of grasses that suggested repeated clearing of the land. Bush concluded that others had been too quick to extrapolate evidence of dense populations in the eastern Amazon across the entire basin. "We don't believe in a virgin system, but we don't believe in a completely exploited system either," he says. "We cannot assume that Amazonian forests were resilient in the face of heavy pre-Columbian disturbance."

McMichael says that she gleaned something similar from Lake Ayauchi. She found significant amounts of charcoal and tiny fossilized structures from maize crops around the lake, but evidence of occupation fell away just a kilometre from the waters[7]. "People were there, but their impact was very localized," she says.

FIERY DEBATE

The paper by McMichael *et al*. was attacked by both sides. Meggers criticized them for accepting existing evidence of large-scale civilization in eastern and central Amazonia. But the strongest criticism came from a dozen scientists who commented on the *Science* paper online. They argued that the paper underestimated signs of civilization by relying too heavily on soil data, which cannot reveal signs of manioc cultivation or other agroforest management. "Soils tell you one part of the story, but they don't tell you everything," says Susanna Hecht, a historical ecologist at the University of California, Los Angeles. She notes that the Inca had established a military outpost in the Putumayo region of western Amazonia by the end of the fifteenth century, just before the arrival of the Europeans, presumably to protect against, or trade with, populations in that region. "Why would you do that if you felt that there were just a bunch of naked people running around?"

In November 2012, a team led by Charles Clement, a horticulturist at the National Institute of Amazonian Research in Manaus, Brazil, countered with a paper[8] arguing that human influence on the forests was pervasive. His study showed that trees that are considered useful for humans — including palms and Brazil nut trees — are more common near rivers, where populations would have been highest. Given that many of these species have different ecological preferences and would not necessarily grow together, the team argued that people



SOURCE: CRYSTAL MCMICHAEL

must have had a hand in creating these bounteous patches. But their results also came under fire. Critics questioned both the strength of the statistics and the fact that the team did not attempt to establish a baseline for what might be expected in the absence of humans.

For researchers on either side of the debate, part of the problem is the sheer scale of the Amazon basin and the difficulty in reaching remote study sites. To get a clearer image, researchers are focusing on the view from space.

After her time in the field with Bush, McMichael joined a team of remote-sensing scientists at the University of New Hampshire in Durham in an attempt to identify *terra preta* through satellite imagery. Their unpublished analysis suggests that it is possible to spot *terra preta* from space by seeking fertile regions where leaves absorb high amounts of nutrients and water.

McMichael has also developed a model that uses ecological factors such as distance from a river, elevation and forest type to predict where *terra preta* might be expected, which has helped field scientists to narrow their search. When tested against 2,900 sites where *terra preta* has been confirmed as either present or absent, the model predicts 89% of them correctly. It suggests that about 3% or 4% of the basin might yield *terra preta*, compared with earlier, broader estimates of 0.1–10%. McMichael is now building a similar model to predict the extent of major earthworks such as those found on Marajó and in western Amazonia.

"If you looked at the forest you wouldn't realize there was any ancient disturbance."

PAST AND FUTURE

It might take a much broader view — in time as well as space — to get the full picture of human impact on the Amazon. People were living there long before pre-Columbian times, and could have been shaping and determining the composition of the forest for millennia through hunting and foraging.

The arrival of humans in the Amazon region coincided with — and perhaps contributed to — the widespread extinctions of megafauna, such as giant sloths and a cousin of the modern elephant, towards the end of the last glacial period some 12,000 years ago. A recent study[9] suggests that these extinctions decreased the availability of the important nutrient phosphorous in the forest, and that might explain why the forest is still limited in phosphorous today.

An analysis of pollen and charcoal records led by Francis Mayle at the University of Edinburgh, UK, suggests that dense rainforest might have given way to savannah and more open vegetation in many areas around 4,000–8,000 years ago thanks to a warmer climate and the use of fire by humans[10]. Mayle wrote that humans might have been "the key agents of disturbance" in the early forest.

Hecht's own work has focused on the more recent past: the 1880s to the 1920s, an era in which massive industrialization and globalization drove a boom in rubber extraction in the Amazon. Hecht compiled data on rubber sales and production and then teamed up with Sassan Saatchi, a remote-sensing expert at NASA's Jet Propulsion Laboratory in Pasadena, California, to map out the impact of rubber production across the basin. Their as-yet-unpublished analysis suggests that a million rubber trees were felled annually for 30 years, with each towering giant bringing down perhaps another ten trees as it fell. Saatchi suspects that the forest is still recovering from this sudden pulse of destruction, and he is trying to see whether signs of regrowth can be detected in the satellite data.

If many areas of the Amazon have been regrowing since the early 1900s, that could have temporarily boosted how much carbon dioxide the region is absorbing from the atmosphere, says Saatchi. If true, as that disturbance fades, the forest's role as a carbon sink could shrink.

Determining the full extent of human influence on the forest will take time. But Heckenberger says that what we know already offers some lessons for today. The evidence of ancient *terra preta* and cultivated 'orchards' in the forest demonstrates that indigenous peoples knew how to manage life within the Amazon long before the days of chainsaws and artificial fertilizers. Future efforts to develop the kind of agroforest systems that are being uncovered from the past, says Heckenberger, would help modern communities to both preserve the Amazon and carve out a living. "These are solutions that are quite intelligent and palatable," says Heckenberger — for all of the Amazon's population. "I think they can learn a lot from themselves." ■

Jeff Tollefson *covers climate, energy and the environment for* Nature.

1. McMichael, C. H. *et al. Science* **336,** 1429–1431 (2012).
2. Meggers, B. J. *Am. Anthropol.* **56,** 801–824 (1954).
3. Roosevelt, A. C., Housley, R. A., Imazio da Silveira, M., Maranca, S. & Johnson, R. *Science* **254,** 1621–1624 (1991).
4. Roosevelt, A. C. *Moundbuilders of the Amazon: Geophysical Archaeology on Marajo Island, Brazil* (Academic Press, 1991).
5. Balée, W. *L'Homme* **33,** 231–254 (1993).
6. Heckenberger, M. J. *et al. Science* **301,** 1710–1714 (2003).
7. McMichael, C. H. *et al. Holocene* **22,** 131–141 (2011).
8. Levis, C. *et al. PLoS ONE* **7,** e48559 (2012).
9. Doughty, C. E., Wolf, A. & Malhi, Y. *Nature Geosci.* **6,** 761–764 (2013).
10. Mayle, F. E. & Power, M. J. *Phil. Trans R. Soc. B* **363,** 1829–1838 (2008).

JAMES WILKINSON/SIO-CALCOFI

An instrument of the California Cooperative Oceanic Fisheries Investigations survey profiles ocean conditions off the US west coast.

Follow the fish

A global, long-term programme of ecological monitoring is needed to track ocean health, say **J. Anthony Koslow** and **Jennifer Couture**.

Next year, the United Nations aims to complete its first World Ocean Assessment, a process akin to the regular reporting of the Intergovernmental Panel on Climate Change. The assessment is timely and crucial: the world's oceans are threatened by many anthropogenic stressors, from pollutants, nutrient runoff and overfishing to warming, deoxygenation and acidification[1]. Current ocean-observation programmes are not fit for purpose.

Variables such as temperature, salinity and chlorophyll levels are monitored globally by satellites, water-column-profiling floats and moored sensor arrays. Ecological monitoring of marine systems, by contrast, is woefully inadequate and has been too long dismissed as too hard and too costly. As a result, the phytoplankton, zooplankton and micronekton (krill and small fish) that comprise the bulk of ocean ecosystems are examined on an ad hoc basis, rather than systematically.

Most natural ocean processes vary on annual to decadal timescales. To identify shifts resulting from climate change, observations need to span 50 years or more. Neglect of ecological monitoring has left the ocean-science community effectively bereft of such long-term data. The Census of Marine Life, for example, is an international effort that aims to characterize marine biodiversity, but over only a decade. Today, there are just two ecological data sets that are sufficiently long to fit the bill — and each has limitations.

A global ocean-observation network needs to be established within the next five years to provide baselines against which ocean health can be assessed in the coming century. Alongside physical oceanographic data, such a network must track the status of species in marine ecosystems around the world.

DECADAL MONITORING

Ecological time series are the Cinderellas of ocean science: long-neglected drudges, eking out their existence at the edge of what is in fashion, they now find themselves in favour at the climate change ball.

Their dismal state was brought home last year when one of us (J.A.K.) reported[2] a 63% decline in the abundance of 24 species of mid-water (mesopelagic) fish off the coast of California, apparently in response to reduced oxygen levels at those depths ▶

(200–1,000 metres). The study[2] was based on ecological and oceanographic time-series data from the California Cooperative Oceanic Fisheries Investigations (CalCOFI), which has sampled the waters off southern California since 1951. The results had wide ramifications.

Reduced oxygen levels at mid-ocean depths are a predicted consequence of global warming, because temperature affects ocean stratification, mixing and ventilation[3]. Southern California waters contain a diverse mix fish species, the habitats of which extend to both the cooler northern Pacific and warmer subtropical waters. On setting up a special session at the 2012 annual meeting of the North Pacific Marine Science Organization to examine trends across the northern Pacific, where similar oxygen declines had been reported[4,5], it emerged that there are no other time-series data sets for the mid-water fauna.

Monitoring what happens at mid-ocean depths is crucial for judging marine health. This layer is the metabolic engine of the ocean: 90% of the organic matter that sinks from the surface is processed there. Many more species of fish live at mid-ocean depths than near the surface. The global biomass of mesopelagic fishes is estimated to be 1 billion–10 billion tonnes — more than an order of magnitude greater than that of commercial marine fish landings, which is less than 100 million tonnes[6].

Many fishes reside at mid-water depths in the day to avoid predators, and rise towards the surface to feed at night in the largest animal migration on Earth. As major consumers of zooplankton and prey of higher predators, mid-water fishes are a key vector in the transport of carbon to the deep ocean.

So, why are there so few time series for marine ecosystems? Most ocean observations come from coastal or fishery programmes. Coastal stations record quantities that are easy to measure from a pier (temperature, salinity and chlorophyll), are crucial to water quality (algae, nutrients and oxygen levels) or conditions that are relevant to local shipping (winds, waves and currents). Fishery surveys focus on commercially exploited species. Most oceanographic sampling is limited to variables that can be measured routinely from ships, moorings, gliders and floats.

In the 1990s, the Global Ocean Observing System (GOOS), developed by the UN's Intergovernmental Oceanographic Commission, held out promise for a worldwide monitoring network. But funding has limited its achievements to predominantly physical measurements — the low-hanging fruit. As part of the open-ocean GOOS programme, the Argo project successfully seeded the world's oceans with some 3,000 floats to profile temperature and salinity. But the coastal GOOS programmes focused on easy observations and local hazards.

Few GOOS operations are systematic. The United States, for example, is served by 11 regional associations, each with its own instrumentation. Assessments speak of the lack of ecological observations as a persistent 'gap' in the programme[7]. In November, GOOS will be holding its first Expert Workshop in Townsville, Australia, to establish panels to cover biogeochemistry, and biology and ecosystems. The need is pressing.

Thanks to Argo and satellites, there is now excellent coverage of temperature, salinity and chlorophyll for a good proportion of the world's oceans. But there are still few ecological time series, in particular species-level data for the plankton that form the base of marine food webs.

Plankton can be surveyed using simple technology. Some of the strongest evidence of climate-linked changes in zooplankton abundance has come from the Continuous Plankton Recorder (CPR) survey of the North Sea and north Atlantic, dating to 1946. The rugged CPR instrument (originally brass but now stainless steel) samples the plankton while being towed by merchant ships on regular routes. The plankton are embedded on a silk strip and spooled into a tank of the preservative formaldehyde, then their species identified. The resulting time-series data have proved to be a treasure trove for understanding the ebb and flow of plankton and fish populations in relation to natural climate oscillations in the north Atlantic and to climate warming[8].

"Ad hoc ocean-monitoring arrangements are poorly suited to current needs."

The longevity of the CPR programme is due to its simplicity. CalCOFI, conversely, is a comprehensive project that endeavours to monitor the oceans from 'winds to whales'. As a joint programme between the Scripps Institution of Oceanography in La Jolla, California, and the state and federal agencies responsible for fisheries since 1949, CalCOFI's quarterly acoustic and egg and larval surveys provide data to assess commercial species such as sardine, anchovy and hake.

From the outset, the CalCOFI surveys have been embedded in a suite of observations of the fishes' physical and biological environments. They yield much of our understanding of how the California Current ecosystem responds to drastic fluctuations in temperature and productivity caused by the El Niño cycle, an abnormal warming of ocean waters in the eastern tropical Pacific, as well as to decadal changes in ocean currents and to climate change[9].

As well as CPR and CalCOFI, there are notable programmes being set up or expanded. These include the Japanese Odate project, which analyses zooplankton collections off the coast of Japan for their species composition. But it remains to be seen whether these programmes will run long enough to identify climate-related changes.

FISCALLY FEASIBLE

Today, the ocean-science community is in the embarrassing position of being asked to assess the state of the world's oceans but lacking the means to do so. Yet it is technically and economically feasible to develop multidisciplinary programmes along the lines of CalCOFI that simultaneously meet marine-management, conservation, scientific and infrastructure needs.

The CalCOFI programme costs about US$5 million a year. Instituting similar initiatives globally to cover the roughly 50 large marine ecosystems that characterize the bulk of ocean life[10] would amount to about $250 million a year. If efforts were coordinated across existing fisheries and environmental-monitoring schemes, the extra cost would be lower.

The United States would need only to find $30 million a year to monitor its six large marine ecosystems. This amount approaches the projected $55-million yearly running cost of the US Ocean Observatories Initiative (OOI), the $386-million infrastructure cost of which has been agreed by the US National Science Foundation. From 2015, the OOI will provide a range of physical, chemical, geological and biological observations for a few slices of the ocean over 25 years, but only a few of the species-level observations that are required to assess ecosystem status.

The impending World Ocean Assessment makes clear that the ad hoc ocean-monitoring arrangements developed since the mid-twentieth century are poorly suited to current needs. Ecological baselines need to be established and time series maintained. This essential task is neither too difficult nor too costly. We hope that the upcoming GOOS workshop will mark a watershed. ■

J. Anthony Koslow *and* **Jennifer Couture** *are at the Scripps Institution of Oceanography, University of California, La Jolla, 92093, USA.*
e-mail: jkoslow@ucsd.edu.

1. Roberts, C. *The Ocean of Life: The Fate of Man and the Sea* (Viking Adult, 2012).
2. Koslow, J. A., Goericke, R., Lara-Lopez, A. & Watson, W. *Mar. Ecol. Prog. Ser.* **436,** 207–218 (2011).
3. Shaffer, G., Olsen, S. M. & Pedersen, J. O. P. *Nature Geosci.* **2,** 105–109 (2009).
4. Stramma, L., Johnson, G. C., Sprintall, J. & Mohrholz, V. *Science* **320,** 655–658 (2008).
5. Deutsch, C., Brix, H., Ito, T., Frenzel, H. & Thompson L. *Science* **333,** 336–339 (2011).
6. Kaartvedt, S., Staby, A. & Aksnes, D. L. *Mar. Ecol. Prog. Ser.* **456,** 1–6 (2012).
7. Alverson, K. *J. Ocean Technol.* **3,** 19–23 (2008).
8. Beaugrand, G., Reid, P. C., Ibañez, F., Lindley, J. A. & Edwards, M. *Science* **296,** 1692–1694 (2002).
9. Ohman, M. D. & Venrick, E. L. *Oceanography* **16,** 76–85 (2003).
10. Longhurst, A. R. *Ecological Geography of the Sea* (Academic Press, 1998).

JEFF GILBERT/ALAMY

Controlling the flow of people entering a nation is a key element of migration policy.

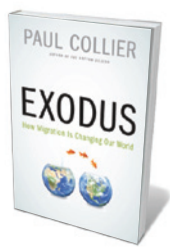
MIGRATION

The porous frontier

L. Alan Winters weighs up a call for evidence-based debate on international migration.

In *Exodus*, Paul Collier's objective is to initiate an evidence-based debate about international migration and debunk the biases — especially among 'liberal thinkers' — that he says make such a debate taboo. Worldwide, about 3% of people live in a country other than that of their birth, but in England and Wales, for example, that proportion is 13% and rising. Collier argues that liberal elites wilfully ignore the views of ordinary citizens, who think that this influx can undermine society and the economy. He provides an analytical framework to account for migration, surveys some of the evidence, identifies supposed intellectual biases and offers policy conclusions.

Collier makes many powerful and thought-provoking points accessibly, and migration scholars would mostly agree with his analysis. But his contribution to the debate is weakened by occasional lapses in consistency, citation and tolerance for opposing views (including mine). For example, Collier argues that the policy question is not whether migration is good or bad, but whether a bit more of it is desirable. However, most economists would argue this way too, and Collier himself periodically slips into extremes. For example, he writes that "the obligation to help the poor cannot imply a generalized obligation to permit free movement of people". Who said that it did? Referring to the migration debate, Collier predicts that the "guardians of orthodoxies stand ready with their fatwas". It would be a better debate if there were no fatwas, either pursued or perceived.



Exodus: How Migration is Changing Our World
PAUL COLLIER
Oxford University Press: 2013.

Considering the host countries, Collier contends that the economic benefits of migration are small and the social costs potentially large. He argues persuasively that when a society has too many incomers, trust is eroded and that this undermines the provision of public goods, including support for society's weaker members. This belief underpins Collier's paean to nationalism as an important, if not vital, cement for societies. Collier justifiably provides a reminder that nationalism need not be militaristic or warmongering.

He also argues that, left unconstrained, migration from poor to rich countries would increase almost without limit, because, as more migrants settle in the rich country, it is easier for others to follow. On this basis, he says that migration must be controlled. Almost all migration scholars recognize this relationship and its implication, so it is wearing that Collier inveighs so frequently against "the open door favored by economists" and "universalist utilitarianism", which he says favours moving "the entire world population … to the country in which people were most productive". Frustratingly, he offers no citation for these views.

International migration moves people from regions of low to high productivity, and it is widely acknowledged that migrants themselves reap almost all the economic benefits through their increase in income. Collier argues that it might be reasonable for a share of these benefits to accrue to host societies, because it is their struggles that ▶

have created the high-productivity environments. But he wisely suggests that trying to collect that share would do more social harm than economic good.

Collier contends that the populations of small poor countries would experience major losses from the emigration of skilled workers if immigration elsewhere went uncontrolled. This is further grist to his mill — he believes that rich countries should impose controls partly for the sake of poor countries, and should also pay compensation to those who move. The brain drain is vigorously debated among specialists: almost all recognize the possibility of such losses, but many argue that poor countries make such an ineffective use of skills that the losses are small. For example, most qualified physicians in such countries serve the urban elite and have almost no impact on the health of the poor.

On policy, Collier recognizes that temporary migration programmes have widespread economic benefits. However, citing the example of Turkish people in Germany, he argues that open liberal democracies cannot enforce departure when temporary migrants' contracts end. He accuses advocates of such temporary mobility (specifically including me) of ignoring non-economic aspects of migration and of having a "tin-eared detachment from a workable ethics".

"The brain drain is vigorously debated among specialists."

His recommendations include requiring the return of asylum seekers when their countries stabilize, and granting "the initial status of guest worker" to all entrants (other than to those who join a lottery for permanent immigrant status). Such guest workers would join a queue to become permanent immigrants, but until they gained that status, they would pay taxes, receive no social benefits and have only limited access to public services. If they declined to register for permanent immigration, they could be deported without appeal. This is not a formal guest-worker scheme, which requires people to leave when their work contracts expire, but is it more ethical?

Collier's book offers a feast of ideas. For this I commend it, but the dominance of rhetorical spice over evidence-based nutrition makes the meal rather indigestible. ■

L. Alan Winters *is professor of economics at the University of Sussex, Brighton, UK; and chief executive of the 'Migrating Out of Poverty' research consortium.*
e-mail: l.a.winters@sussex.ac.uk

A distant HMS *Beagle* off Tierra del Fuego in South America in the 1830s, painted by Conrad Martens.

NATIONAL MARITIME MUSEUM, GREENWICH, LONDON

NATURAL HISTORY

Hell in the Pacific

A turbulent history of early naturalists and the maritime explorers who hosted them fascinates **Andrew Robinson**.

A keen tension runs through Glyn Williams's *Naturalists at Sea*, his chronicle of 14 Pacific Ocean expeditions spanning the seventeenth to nineteenth centuries. The source of this tension is the often-fraught dynamic between the expeditions' on-board naturalists and the seamen who made their research possible. Fieldwork, onerous enough in the primitive naval conditions of the era, was often scuppered by the demands of maritime exploration and surveying, and by international political rivalries.

So Williams's odyssey — beginning with English buccaneer William Dampier's foray to New Holland (modern Australia) in the 1680s, and sailing on through other celebrated English, French, Russian and Spanish voyages — is almost as much a history of psychology as of scientific derring-do and discovery. Even Captain James Cook, who on his first great voyage had harmoniously hosted naturalist Joseph Banks, was not immune to discord. On his second expedition, in 1772–75, Cook fell out with naturalist Johann Reinhold Forster, and before setting off on his third, with the ship's surgeon now doubling as a naturalist, is said to have exclaimed: "Curse scientists, and all science into the bargain."

During the 1831–36 circumnavigation of HMS *Beagle*, the conservative Captain Robert FitzRoy famously clashed with the young Charles Darwin over slavery in Brazil; Darwin even considered leaving the ship. Dampier's expedition avoided such tension, because Dampier was both a seaman and a naturalist. Among his many beguiling descriptions is that of the hummingbird, which "haunts about Flowers and Fruit, like a Bee gathering Honey, making many near addresses to its delightful Objects".

➲ NATURE.COM
For more on Darwin and HMS *Beagle*:
nature.com/darwin

The main cause of all this on-board strain, argues Williams, was identified by biologist Thomas Henry Huxley. In the 1840s, Huxley served as an assistant surgeon in the Royal Navy before becoming a celebrated scientist. The hard physical work of the sailor, "in his constant battle with the elements, is as far apart from the speculative acuteness and abstraction necessary to the man of science as ever", he wrote in an 1854 essay.

His friend Darwin had a more personal explanation. After four years away from home, he wrote from Tasmania: "I hate every wave of the ocean … I believe there are very few contented Sailors. — They are caught young & broken in before they have reached years of discretion. Those who are employed, sigh after the delights of the shore, & those on shore, complain they are forgotten & overlooked."

As Williams explains, whatever their psychology, naval captains of these times were concerned first and foremost with the safety of their ships. Their second priority was to explore and survey coastlines, while claiming lands for their home countries and searching for the fabled Northwest Passage or *Terra Australis Incognita* — the unknown southern continent of Antarctica. Finally, they had to consider the vagaries of European politics and wars, which could see the imprisonment of an expedition's members. This happened, for instance, when Joseph-Antoine Bruny d'Entrecasteaux's expedition, dispatched from France in 1791, docked its two vessels in

Naturalists at Sea: Scientific Travellers from Dampier to Darwin
GLYN WILLIAMS
Yale University Press: 2013.

the Dutch East Indies in late 1793. After the crew learned with shock that Louis XVI had been executed and that the French Republic was at war with the Dutch, most of them were incarcerated and the two ships were seized.

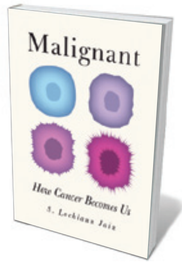
Naturalists, by contrast, cared most about having enough time to go ashore and discover new flora and fauna. After collecting specimens, often with great difficulty, they depended on the captain's goodwill to enable them to draw and record their trophies in cramped and inclement conditions and, if possible, to keep them alive for the journey home. Darwin was fortunate to be able to spend 60% of his voyage time on land and to dispatch specimens regularly from South America to Britain. Some 70% of Cook's second voyage, however, was spent at sea — to the despair of Forster, who wrote that "after having circumnavigated very near half the globe we saw nothing, but water, Ice & Sky".

Conflict was inevitable, Williams concludes from studying such records. On Vitus Bering's 1741 Russian expedition to the strait between Russia and North America that now bears his name, naturalist Georg Wilhelm Steller had less than a day in newly discovered Alaska. He managed to walk several miles along a beach collecting plants, but when he asked Bering for a small boat and several men, was told to return within an hour or be left behind. In his journal, Steller complained that "the preparation for this ultimate purpose lasted ten years; twenty hours were devoted to the matter itself". On George Vancouver's British expedition to the Pacific in 1791–95, an on-board greenhouse full of specimen plants, erected on the instruction of Banks (then president of the Royal Society), became a bone of contention. The burden for the crew of keeping the plants alive eventually led Vancouver to arrest the ship's naturalist, Archibald Menzies, for insubordination.

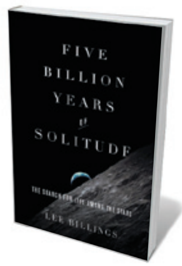
Williams has been researching the history of European incursions into the Pacific and Arctic oceans since the late 1950s, and has published many books on the subject. An erudite and beautifully illustrated work, *Naturalists at Sea* wears its learning lightly, and conveys to non-specialists an array of fascinating details about explorers and naturalists, familiar and not-so-familiar, quoting judiciously from their journals and post-voyage publications. Although it sometimes struggles to bring its dizzying cast of characters to life, every page testifies to the indomitable vitality of both explorers and naturalists. ■

Andrew Robinson *is author of* The Story of Measurement *and* The Shape of the World: The Mapping and Discovery of the Earth.
e-mail: andrew.robinson33@virgin.net

Books in brief

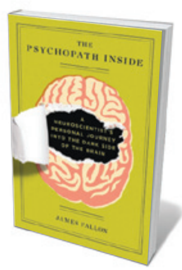


Malignant: How Cancer Becomes Us
S. Lochlann Jain UNIVERSITY OF CALIFORNIA PRESS *(2013)*
Patients with cancer generate so much revenue for the US health-care industry that a cure would be an economic risk. Thus argues anthropologist S. Lochlann Jain, who deems cancer "a constitutive aspect of American social life, economics, and science" — so bizarrely entwined that chemical companies churn out both cancer drugs and carcinogenic herbicides. In this trenchant mix of science history, memoir and cultural analysis, Jain is thoughtful and often darkly humorous on everything from cancer statistics to treatments, trials and issues around sexuality. Brilliant and disturbing.

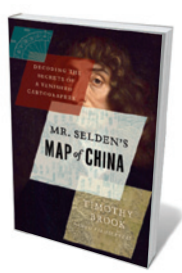


Five Billion Years of Solitude: The Search for Life Among the Stars
Lee Billings CURRENT *(2013)*
Unsurprisingly for an infant science, the quest for other Earths is sometimes fractious. Science writer Lee Billings deftly captures both behind-the-scenes ructions and landmark discoveries in his tour of this multidisciplinary field, its history and its players. The seamlessly interwoven narrative is strong on big personalities, from astronomer Frank Drake, a pioneer of the search for extraterrestrial intelligence (SETI) whose work is now overshadowed by glamorous finds in exoplanetary science, to astrophysicist Sara Seager, a scintillating star in that very field.

High Moon Over the Amazon: My Quest to Understand the Monkeys of the Night
Patricia Chapple Wright LANTERN *(2013)*
One-time "hippie housewife" Patricia Chapple Wright became a trailblazing primatologist by dint of determination and sheer curiosity. As related in this engaging memoir, her scientific odyssey began in a New York pet shop in the 1960s. After purchasing an owl monkey (*Aotus lemurinus griseimembra*), she travelled to Peru to locate a mate for it — and recognized her métier. Decades on, now a world authority on lemurs, she has set up the Ranomafana National Park and an adjacent research site, Centre ValBio, in Madagascar.



The Psychopath Inside: A Neuroscientist's Personal Journey into the Dark Side of the Brain
James Fallon CURRENT *(2013)*
In 2005, neuroscientist James Fallon was checking the brain scans of psychopathic murderers and 'normal' controls, including himself. Noting that his scan closely resembled those of the murderers, the happy, successful Fallon had to know why. He shares his journey, mining genetics, epigenetics and neuroscience, and perusing his childhood (including a brief spell of obsessive–compulsive disorder), family tree and behavioural eccentricities. His surprising final diagnosis could broaden the way we see normality.



Mr. Selden's Map of China: Decoding the Secrets of a Vanished Cartographer
Timothy Brook BLOOMSBURY *(2013)*
The Selden map of China and its environs, an anonymous cartographic puzzle unearthed in the Bodleian Library in 2009, is the pivot for this cultural history. Timothy Brook illuminates the map's odd features and backstory. Along with the lives of those tangled in its history (such as Michael Shen, the Chinese Jesuit who translated the map's script in the late 1600s), Brook reveals how the amazingly accurate chart hints at the first stirrings of globalization. **Barbara Kiser**

have created the high-productivity environments. But he wisely suggests that trying to collect that share would do more social harm than economic good.

Collier contends that the populations of small poor countries would experience major losses from the emigration of skilled workers if immigration elsewhere went uncontrolled. This is further grist to his mill — he believes that rich countries should impose controls partly for the sake of poor countries, and should also pay compensation to those who move. The brain drain is vigorously debated among specialists: almost all recognize the possibility of such losses, but many argue that poor countries make such an ineffective use of skills that the losses are small. For example, most qualified physicians in such countries serve the urban elite and have almost no impact on the health of the poor.

On policy, Collier recognizes that temporary migration programmes have widespread economic benefits. However, citing the example of Turkish people in Germany, he argues that open liberal democracies cannot enforce departure when temporary migrants' contracts end. He accuses advocates of such temporary mobility (specifically including me) of ignoring non-economic aspects of migration and of having a "tin-eared detachment from a workable ethics".

> ***"The brain drain is vigorously debated among specialists."***

His recommendations include requiring the return of asylum seekers when their countries stabilize, and granting "the initial status of guest worker" to all entrants (other than to those who join a lottery for permanent immigrant status). Such guest workers would join a queue to become permanent immigrants, but until they gained that status, they would pay taxes, receive no social benefits and have only limited access to public services. If they declined to register for permanent immigration, they could be deported without appeal. This is not a formal guest-worker scheme, which requires people to leave when their work contracts expire, but is it more ethical?

Collier's book offers a feast of ideas. For this I commend it, but the dominance of rhetorical spice over evidence-based nutrition makes the meal rather indigestible. ■

L. Alan Winters *is professor of economics at the University of Sussex, Brighton, UK; and chief executive of the 'Migrating Out of Poverty' research consortium.*
e-mail: l.a.winters@sussex.ac.uk

NATIONAL MARITIME MUSEUM, GREENWICH, LONDON

A distant HMS *Beagle* off Tierra del Fuego in South America in the 1830s, painted by Conrad Martens.

NATURAL HISTORY

Hell in the Pacific

A turbulent history of early naturalists and the maritime explorers who hosted them fascinates **Andrew Robinson**.

A keen tension runs through Glyn Williams's *Naturalists at Sea*, his chronicle of 14 Pacific Ocean expeditions spanning the seventeenth to nineteenth centuries. The source of this tension is the often-fraught dynamic between the expeditions' on-board naturalists and the seamen who made their research possible. Fieldwork, onerous enough in the primitive naval conditions of the era, was often scuppered by the demands of maritime exploration and surveying, and by international political rivalries.

So Williams's odyssey — beginning with English buccaneer William Dampier's foray to New Holland (modern Australia) in the 1680s, and sailing on through other celebrated English, French, Russian and Spanish voyages — is almost as much a history of psychology as of scientific derring-do and discovery. Even Captain James Cook, who on his first great voyage had harmoniously hosted naturalist Joseph Banks, was not immune to discord. On his second expedition, in 1772–75, Cook fell out with naturalist Johann Reinhold Forster, and before setting off on his third, with the ship's surgeon now doubling as a naturalist, is said to have exclaimed: "Curse scientists, and all science into the bargain."

During the 1831–36 circumnavigation of HMS *Beagle*, the conservative Captain Robert FitzRoy famously clashed with the young Charles Darwin over slavery in Brazil; Darwin even considered leaving the ship. Dampier's expedition avoided such tension, because Dampier was both a seaman and a naturalist. Among his many beguiling descriptions is that of the hummingbird, which "haunts about Flowers and Fruit, like a Bee gathering Honey, making many near addresses to its delightful Objects".

> **➲ NATURE.COM**
> For more on Darwin and HMS *Beagle*:
> nature.com/darwin

The main cause of all this on-board strain, argues Williams, was identified by biologist Thomas Henry Huxley. In the 1840s, Huxley served as an assistant surgeon in the Royal Navy before becoming a celebrated scientist. The hard physical work of the sailor, "in his constant battle with the elements, is as far apart from the speculative acuteness and abstraction necessary to the man of science as ever", he wrote in an 1854 essay.

His friend Darwin had a more personal explanation. After four years away from home, he wrote from Tasmania: "I hate every wave of the ocean … I believe there are very few contented Sailors. — They are caught young & broken in before they have reached years of discretion. Those who are employed, sigh after the delights of the shore, & those on shore, complain they are forgotten & overlooked."

As Williams explains, whatever their psychology, naval captains of these times were concerned first and foremost with the safety of their ships. Their second priority was to explore and survey coastlines, while claiming lands for their home countries and searching for the fabled Northwest Passage or *Terra Australis Incognita* — the unknown southern continent of Antarctica. Finally, they had to consider the vagaries of European politics and wars, which could see the imprisonment of an expedition's members. This happened, for instance, when Joseph-Antoine Bruny d'Entrecasteaux's expedition, dispatched from France in 1791, docked its two vessels in

Naturalists at Sea: Scientific Travellers from Dampier to Darwin
GLYN WILLIAMS
Yale University Press: 2013.

the Dutch East Indies in late 1793. After the crew learned with shock that Louis XVI had been executed and that the French Republic was at war with the Dutch, most of them were incarcerated and the two ships were seized.

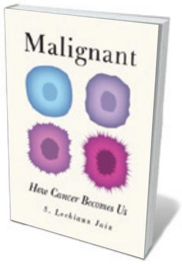
Naturalists, by contrast, cared most about having enough time to go ashore and discover new flora and fauna. After collecting specimens, often with great difficulty, they depended on the captain's goodwill to enable them to draw and record their trophies in cramped and inclement conditions and, if possible, to keep them alive for the journey home. Darwin was fortunate to be able to spend 60% of his voyage time on land and to dispatch specimens regularly from South America to Britain. Some 70% of Cook's second voyage, however, was spent at sea — to the despair of Forster, who wrote that "after having circumnavigated very near half the globe we saw nothing, but water, Ice & Sky".

Conflict was inevitable, Williams concludes from studying such records. On Vitus Bering's 1741 Russian expedition to the strait between Russia and North America that now bears his name, naturalist Georg Wilhelm Steller had less than a day in newly discovered Alaska. He managed to walk several miles along a beach collecting plants, but when he asked Bering for a small boat and several men, was told to return within an hour or be left behind. In his journal, Steller complained that "the preparation for this ultimate purpose lasted ten years; twenty hours were devoted to the matter itself". On George Vancouver's British expedition to the Pacific in 1791–95, an on-board greenhouse full of specimen plants, erected on the instruction of Banks (then president of the Royal Society), became a bone of contention. The burden for the crew of keeping the plants alive eventually led Vancouver to arrest the ship's naturalist, Archibald Menzies, for insubordination.

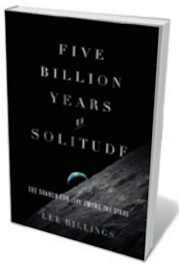
Williams has been researching the history of European incursions into the Pacific and Arctic oceans since the late 1950s, and has published many books on the subject. An erudite and beautifully illustrated work, *Naturalists at Sea* wears its learning lightly, and conveys to non-specialists an array of fascinating details about explorers and naturalists, familiar and not-so-familiar, quoting judiciously from their journals and post-voyage publications. Although it sometimes struggles to bring its dizzying cast of characters to life, every page testifies to the indomitable vitality of both explorers and naturalists. ■

Andrew Robinson *is author of* The Story of Measurement *and* The Shape of the World: The Mapping and Discovery of the Earth.
e-mail: andrew.robinson33@virgin.net

Books in brief

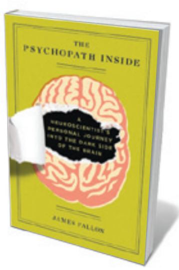


Malignant: How Cancer Becomes Us
S. Lochlann Jain UNIVERSITY OF CALIFORNIA PRESS *(2013)*
Patients with cancer generate so much revenue for the US health-care industry that a cure would be an economic risk. Thus argues anthropologist S. Lochlann Jain, who deems cancer "a constitutive aspect of American social life, economics, and science" — so bizarrely entwined that chemical companies churn out both cancer drugs and carcinogenic herbicides. In this trenchant mix of science history, memoir and cultural analysis, Jain is thoughtful and often darkly humorous on everything from cancer statistics to treatments, trials and issues around sexuality. Brilliant and disturbing.

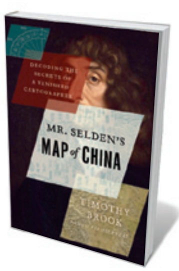


Five Billion Years of Solitude: The Search for Life Among the Stars
Lee Billings CURRENT *(2013)*
Unsurprisingly for an infant science, the quest for other Earths is sometimes fractious. Science writer Lee Billings deftly captures both behind-the-scenes ructions and landmark discoveries in his tour of this multidisciplinary field, its history and its players. The seamlessly interwoven narrative is strong on big personalities, from astronomer Frank Drake, a pioneer of the search for extraterrestrial intelligence (SETI) whose work is now overshadowed by glamorous finds in exoplanetary science, to astrophysicist Sara Seager, a scintillating star in that very field.

High Moon Over the Amazon: My Quest to Understand the Monkeys of the Night
Patricia Chapple Wright LANTERN *(2013)*
One-time "hippie housewife" Patricia Chapple Wright became a trailblazing primatologist by dint of determination and sheer curiosity. As related in this engaging memoir, her scientific odyssey began in a New York pet shop in the 1960s. After purchasing an owl monkey (*Aotus lemurinus griseimembra*), she travelled to Peru to locate a mate for it — and recognized her métier. Decades on, now a world authority on lemurs, she has set up the Ranomafana National Park and an adjacent research site, Centre ValBio, in Madagascar.



The Psychopath Inside: A Neuroscientist's Personal Journey into the Dark Side of the Brain
James Fallon CURRENT *(2013)*
In 2005, neuroscientist James Fallon was checking the brain scans of psychopathic murderers and 'normal' controls, including himself. Noting that his scan closely resembled those of the murderers, the happy, successful Fallon had to know why. He shares his journey, mining genetics, epigenetics and neuroscience, and perusing his childhood (including a brief spell of obsessive–compulsive disorder), family tree and behavioural eccentricities. His surprising final diagnosis could broaden the way we see normality.



Mr. Selden's Map of China: Decoding the Secrets of a Vanished Cartographer
Timothy Brook BLOOMSBURY *(2013)*
The Selden map of China and its environs, an anonymous cartographic puzzle unearthed in the Bodleian Library in 2009, is the pivot for this cultural history. Timothy Brook illuminates the map's odd features and backstory. Along with the lives of those tangled in its history (such as Michael Shen, the Chinese Jesuit who translated the map's script in the late 1600s), Brook reveals how the amazingly accurate chart hints at the first stirrings of globalization. **Barbara Kiser**

Members of Random Dance in *Becoming*, a work involving a digitized partner seen through 3D glasses.

DAVID BICKERSTAFF

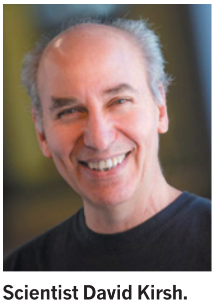
COGNITIVE SCIENCE

Leap of thought

Ewen Callaway meets cognitive scientist David Kirsh, who works with choreographer Wayne McGregor.

Sadler's Wells Theatre, London's leading venue for contemporary dance, is a second home to cognitive scientist David Kirsh. On a muggy September day, young dancers loitering between rehearsals greet him with hellos and hugs. Anna Nowak — a member of Wayne McGregor Random Dance, a company known for its highly kinetic, technologically inspired productions — dashes over to tell Kirsh she loved *Thinking With the Body*. This exhibition, at the Wellcome Collection in London, explores a decade of collaborations between McGregor's company and scientists such as Kirsh.

A visitor could walk through the small show in minutes. Yet *Thinking With the Body* rewards a longer look — particularly the documentary-style videos that show the deep nature of the company's collaborations. Unlike in many science–art partnerships, in which the intersection between the two camps can be tenuous at best, McGregor and his partners all gain much more than a line in their CVs. "I think what is vital in genuine collaboration between art and science is the notion that neither is in the service of the other," McGregor said in 2009.



Scientist David Kirsh.

Thinking With the Body: Mind and Movement in the Work of Wayne McGregor Random Dance
Wellcome Collection, London. Until 27 October 2013.

The most fruitful of these partnerships seems to be the one with Kirsh, who works on embodied cognition — how the body influences thought. Kirsh, who is associate director of the Arthur C. Clarke Center for Human Imagination at the University of California, San Diego, started studying embodied cognition in McGregor's company after observing the choreographer's creative process in London five years ago. He was gobsmacked. "There were dozens of phenomena just dripping in this two-hour display, which would clearly bear careful study," says Kirsh. He was particularly interested in the instructions that McGregor gives to his dancers, which often involve cognitive tasks such as asking someone to move in response to a strap on another dancer's clothing. "These are techniques that give the dancers material they can think about that will generate movement forms that perhaps have never been done before," says Kirsh. McGregor is known for bizarre-sounding instructions to dancers (such as "dance like a skyscraper") to spur them to invent movements that might be used in performances.

As with any scientific project, Kirsh began studying Random Dance by characterizing the different phenomena he saw "like a botanist", he says. "I go out with six or seven high-definition video cameras, I put them around the studio floor and collect everything from the moment he introduces a dancer to the premiere weeks later." He and a trained team of students then deciphered the different techniques for instruction and practice that they saw in the videos, in much the same way that primatologists characterize behaviour.

One phenomenon that caught Kirsh's attention was 'marking', in which dancers in rehearsal elaborate only the basics of a dance movement. "It's a lower-energy version; they won't stretch as far; they won't have the emotional force in it. It's a way to avoid injury and because you can't dance for five hours after two hours of exercises warming up," Kirsh says.

But as he discovered when conducting a controlled experiment, there is more to it. He showed dancers a new routine, gave them time to learn the moves, and divided them into three groups to practise again. One group performed the full movements, a second marked them, and a third lay down and imagined themselves performing the dance. To Kirsh's surprise, the dancers who marked the routine executed it most faithfully later. "Nobody predicted this," he says. "This is the hint at a theory of practising, and now it's open to study this much more carefully to understand how people focus on aspects of what they're practising." The experiment, he feels, is evidence of physical activity influencing thought.

On the basis of his work with Random Dance, Kirsh has published research papers on interaction design, McGregor's creative process and a phenomenon that he calls distributed memory, in which dancers remember dozens of complicated movements through physical cues from other dancers. McGregor, too, has gained from their collaboration. When he instructs dancers and other young choreographers, he now uses terms that Kirsh devised, such as 'sonifications' — sounds that choreographers make to guide how a dancer shapes a move, such as "yah ooh ehh". Kirsh notes, "Now that the term has been named, the phenomenon is clear." Helping Kirsh to understand choreography has made McGregor more self-aware of his own creative process. Kirsh also studied how the dancers interacted with computer-generated forms during rehearsals for *Becoming*.

Other scientists who hope to get as much out of their partnerships with artists could begin by adopting the perspective of their study subjects to understand them and their work, Kirsh says. "The way I came in was partly as an ethnographer, somebody who tries to understand that particular culture as an insider rather than as an outsider," he says. "The more deeply you understand that, the better your discourse with that artist is." ■

WAYNE MCGREGOR RANDOM DANCE

Ewen Callaway *is a reporter for* Nature.

Correspondence

Reduce, reuse and recycle lab waste

Implementing ecological awareness at the bench has saved up to 40% of my research funding over one year. As scientists and good citizens, many of us strive to reduce, reuse and recycle waste at home. We should be doing the same in the laboratory.

For speed and convenience, molecular biologists in particular rely on disposable tools. Nonetheless, a substantial proportion of plastic lab supplies, including pipette-tip boxes, can be reused after washing and sterilizing. Plastic that is free of hazardous material and not reusable can be recycled.

When possible, we should be using autoclavable glassware instead of plastic. This is fine for reagent vessels, tubing and pipettes, for example, and suitable for most applications, including bacterial and tissue culture, and sampling and preparing solutions or aliquots. Glass can also be treated to remove traces of heat-resistant nuclease enzymes that sabotage experiments with nucleic acids.

Asking for more sustainable products, less packaging and more recyclable plastics will also help the lab-supply industry to prosper.

Scientists are educators. We should improve environmental awareness, responsibility and training in our labs to ensure we are not among the last to jump on the sustainability bandwagon.

Gaia Bistulfi *D'Youville College, Buffalo, New York, USA.*
bistulfi@dyc.edu

Understanding our destructive choices

In his review of my book *Invisible Nature: Healing the Destructive Divide Between People and the Environment*, Edward Humes suggests that I should have expanded my psychology argument and given less attention to "long-dead philosophers" and historical contributions (*Nature* **500,** 26–27; 2013). I disagree.

An interdisciplinary approach to environmental issues stands a better chance of explaining our apparently intractable modern destructiveness. And many of modernity's destructive disconnections stem from philosophical assumptions made by the founders of modern science, which can be remedied by alternative theories from dissenting scientists such as the late David Bohm, a theoretical physicist.

Humes also recommends interviewing more people about their harmful choices. But it scarcely makes sense to ask people with no experience of the destructive outcomes of their choices how that lack of experience is not influencing their choices. Furthermore, sampling interviewees would challenge the generality of the study, and would be undermined by the unreliability of self-reported attributions of behaviour.

Instead of attempting to discern good players from bad ones, I believe that it is more fruitful to study the conditions that perpetuate destructive choices by us all.

Kenneth Worthy *University of California, Santa Cruz, USA.*
kworthy@ucsc.edu

Biologists borrow more than words

Eleonore Pauwels is right that biologists should not mislead themselves or the public by using simplistic metaphors borrowed from engineering (*Nature* **500,** 523–524; 2013). But biologists don't simply borrow words, they take engineering principles — derived from theory and practice — and apply them to biological systems.

Referring to the heart as a pump or to the nasal passages as heat exchangers is not a simple metaphor like calling Juliet the Sun. We use the same terms because the same formal criteria can be usefully applied to both engineered and evolved systems.

Shared engineering terminology extends beyond biomechanics to molecular and systems biology. For example, genes have been successfully modelled as Boolean logic switches to predict gene expression in the developing embryo (I. S. Peter *et al. Proc. Natl Acad. Sci. USA* **109,** 16434–16442; 2012).

In my view, it is an oversimplification to think that all engineering talk in biology is mere imagery.

Brett Calcott *Australian National University, Canberra, Australia.*
brett.calcott@gmail.com

Oil-palm replanting raises ecology issues

More than one-third of the area on which oil palm is grown in Malaysia, some 1.4 million hectares (http://faostat.fao.org), has already passed peak yields and is due to be replanted. Replanting, which represents a new phase for the industry, must be carefully thought through and implemented to avoid repeating the disastrous effects of the initial clearance of primary forest on

biodiversity and the environment.

The productive life of an oil-palm crop is 25–30 years and, because the boom in oil-palm cultivation began in the mid-1980s, large areas of ageing oil palm in southeast Asia now need replacing. However, the long-term nature of the crop has allowed biological complexity to build up around it over time.

There is therefore a risk that the replanting phase — just like the initial clearance for planting — will disrupt natural habitats by indiscriminate removal of vegetation and heavy disturbance of soil and hydrological systems.

We suggest that large-scale replanting operations should be carried out carefully with a view to minimizing such adverse effects on the environment.

Jake L. Snaddon, Katherine J. Willis *Biodiversity Institute, University of Oxford, UK.*
jlsnaddon@gmail.com
David W. Macdonald *Wildlife Conservation Research Unit, University of Oxford, UK.*

Fees could damage public data archives

Dryad, one of the largest public data repositories in the life sciences, last month imposed modest fees on researchers for archiving their data (see http://datadryad.org/pages/pricing). We believe that this strategy is misguided. It will further discourage researchers from participating in public data archiving.

Authors' participation in archiving is already low, and many fail to comply with journals' policies in this regard (see, for example, B. T. Drew *Nature* **493,** 305; 2013). Even when their work is publicly funded, researchers can be reluctant to share their data through public archives because they lose priority access to it. And compensatory benefits — such as those arising from increased citation of their papers — are not always forthcoming. Requiring scientists to pay fees from their own research funds is adding one more hurdle.

Of course, public repositories of data need to be maintained and paid for. But rather than billing researchers, costs could be met using more subtle methods — much as they are for 'free' restaurant bread and hotel Internet access. Payment could come from private- or public-sector grants, institutional contributions, dedicated funds quarantined by grant agencies, or from partnerships with journal publishers. That way, funders' budgets or publishers' profits would foot the bill.

Dominique G. Roche, Michael D. Jennions, Sandra A. Binning *Australian National University, Canberra, Australia.*
dominique.roche@anu.edu.au

Celebrating 80 years of science solidarity

The Council for Assisting Refugee Academics (CARA), which supports persecuted academics worldwide, celebrates its 80th anniversary this year. Today, as never before, intellectual leaders in every country need the freedom to teach, research and publish the global solutions on which future generations will depend.

Speaking in October 1933 on behalf of CARA, then known as the Academic Assistance Council, Albert Einstein urged a packed audience at London's Royal Albert Hall to "resist the powers which threaten to suppress intellectual and individual freedom". This was the first major UK fund-raising event to support German academics, many of them Jewish, to whom Adolf Hitler had begun to close German universities.

Among the many helped by CARA during the ensuing years was the German neurosurgeon Ludwig Guttmann. He set up the Stoke Mandeville Spinal Injuries Unit near Aylesbury, UK, where he revolutionized the treatment of people with spinal-cord injuries, for whom palliative care had previously been the only option. In 1948 he founded the Stoke Mandeville Games, which were retrospectively adopted by the International Olympic Committee as the Paralympic Games in 1984.

Today, CARA runs regional programmes focusing on Iraq, Syria and Zimbabwe. In parallel with its core rescue efforts to provide sanctuary to academics at risk, CARA works to sustain those who choose to remain in crisis-affected countries.

Recent achievements include establishing a virtual lecture hall at the University of Zimbabwe in Harare in 2012 to revive the teaching of health and veterinary sciences. Our work is supported by the CARA Scholars at Risk UK Universities Network, now incorporating 95 universities.

To mark our anniversary, we have established an annual 'Einstein Lecture' to explore the link between science and civilization in a modern context (see www.cara1933.org).

Anne Lonsdale *CARA, London, UK.*
info.cara@lsbu.ac.uk

Analyse impact of health priorities

There is a dearth of evidence for the assumed beneficial effects of involving patients and the public in health-research decisions (S. Petit-Zeman and L. Locock *Nature* **501,** 160–161; 2013). This is just one example of the general difficulty of measuring the impact of setting health-research priorities.

These priorities are decided by researchers and policy-makers to stimulate urgent research (see, for example, go.nature.com/pxealf). The guidelines for evaluating such priority-setting processes (R. F. Viergever *et al. Health Res. Policy Sys.* **8,** 36; 2010) need to be extended to cover impact measurement and analysis.

Impact measurements might include: evaluating general changes, such as in the attitudes of researchers and policy-makers; monitoring the research that is undertaken in line with the established priorities (for example, by looking at conducted clinical trials or published articles: see J-A. Røttingen *et al. Lancet* http://doi.org/n4c; 2013); and assessing the impact of that research on factors such as health or socio-economic status (see R. Banzi *et al. Health Res. Policy Sys.* **9,** 26; 2011).

Analysis of these impact measurements could be useful for building evidence to justify and improve prioritization in health research and to monitor progress for informing future priority-setting.

Roderik F. Viergever *Radboud University Nijmegen, the Netherlands; and London School of Hygiene and Tropical Medicine, UK.*
rikviergever@gmail.com

A call to forge biodiversity links

For the upcoming calls for Horizon 2020 research funding, the European Commission has said that it would prefer bids from open, collaborative consortia rather than the competitive bids seen in previous funding programmes. To this end, the organizers of 18 European biodiversity informatics projects agreed at a meeting in Rome last month (see go.nature.com/jouct9) to the overarching goal of predictively modelling the biosphere (see D. Purves *et al. Nature* **493,** 295–297; 2013).

This goal is arguably more complicated than climate modelling, for example. Working out the first steps to take during the Horizon 2020 funding period will hinge on researchers forging links among a wide range of disciplines, including ecology, microbiology, agriculture, socioeconomics, remote sensing, taxonomy and molecular biology.

We therefore invite readers to contribute to the discussion on project ideas at http://h2020.myspecies.info, where consortia are now being built.

Alex Hardisty* *Biodiversity Virtual e-Laboratory (BioVeL), Cardiff University, UK.*
hardistyar@cardiff.ac.uk
**On behalf of 19 co-authors (see go.nature.com/2tqsqc for full list).*

CONTRIBUTIONS

Correspondence may be sent to **correspondence@nature.com** after consulting the guidelines at **http://go.nature.com/cmchno**. Alternatively, readers may comment online: **www.nature.com/nature**.

Peter Huttenlocher

(1931–2013)

Neurologist who showed formation and pruning of synapses in child development.


JAMES BALLARD/UNIV. CHICAGO MEDICINE

Peter Richard Huttenlocher, a wonderfully kind and quietly driven neurologist, showed that billions of synapses — the connections between brain cells that mediate learning and memory — are generated in the human cerebral cortex during the first months of life. He also made the more surprising discovery that many of these synapses are eliminated during subsequent years, just when children achieve the most important developmental milestones, such as walking and speaking.

This idea — that the elimination, or 'pruning', of synapses is as much a part of learning as their formation — has come to influence fields as diverse as developmental neuroscience, child development, early education and native language learning.

Huttenlocher, who died of pneumonia on 15 August after a long struggle with Parkinson's disease, was born in Oberlahnstein bei Koblenz, Germany, in 1931. His mother, an opera singer, refused to join the Nazi party and fled to the United States in 1937, leaving the young Peter and his brothers to be raised by their father, a chemist. Huttenlocher's experiences as a child witness of Nazism during the Second World War, and of the starvation and suffering of German survivors in the Russian and French zones in post-war Germany, contributed to his lifelong interest in ethics, morals and human behaviour.

In 1949, aged 18, he travelled to the United States with his older brother Dieter to visit his mother, and decided to stay. Soon after Peter arrived, he attended the University of Buffalo in New York where he met his future wife, Janellen Burns, who would go on to become a successful cognitive psychologist.

Peter graduated *summa cum laude* in philosophy in 1953, and he and Janellen married a year later. Together, they moved to Harvard University in Massachusetts, where she obtained her PhD in Psychology, and he his doctor of medicine at Harvard Medical School, graduating *magna cum laude* in 1957. After medical residency training and two research fellowships, Huttenlocher spent a couple of years as an instructor at Harvard, followed by eight years as an assistant and then associate professor of paediatrics and neurology at the Yale School of Medicine in New Haven, Connecticut. In 1974, he and Janellen moved to the University of Chicago in Illinois, where he remained for the rest of his career. There, he was a professor of paediatrics and neurology and the founding head of child neurology.

In the mid-1970s, Huttenlocher began to study the formation of synapses in the brains of healthy children and young adults who had died from various causes. Using an electron microscope, and doing much of the work himself, he mapped the developmental trajectory of synaptogenesis in the cerebral cortex from around the time of birth to early adulthood. He originally planned to compare the brains of children with intellectual disabilities to those without, but, as he later wrote in an essay on the subject, he soon discovered that "the findings in the normal population were more interesting than the abnormal population".

Huttenlocher found that the number of synapses increases more than tenfold during the first year after birth. This made sense, given the need for synaptic connections to form the neuronal circuits that underlie new abilities and memories in the first year of a child's life. Surprisingly, however, sometime during the second year of life, the total number of synapses drops drastically. Huttenlocher showed that synaptic pruning continues over several years — just when children are acquiring language, learning to run and starting school, with total synaptic counts stabilizing in adolescence.

Huttenlocher's discovery that synapses are overproduced and then pruned was 20 years or more ahead of its time. Today, most ideas about human brain development, from the microscopic to the macroscopic to the societal, draw on his work. For instance, researchers are now investigating the mechanisms that control pruning, the possibilities that this synaptic plasticity provides, and how to use an understanding of synaptic development to optimize early educational intervention, language learning or music instruction. Moreover, in the past decade, most models of autism spectrum disorder have suggested that people with autism might not have fewer synapses in their brains than people without, but, instead, that they show defects in the removal of synapses that are usually eliminated.

Huttenlocher continued his work on synapses throughout the 1970s. He was among the first to observe that the brains of individuals with intellectual disability have the same number of synapses as those without, but that the synapses tend to be different in shape. This discovery of the relationship between synaptic shape and function is another key element in our understanding of brain plasticity and learning. He also made numerous other contributions; for example, he discovered that an inherited form of epilepsy is associated with a mild brain malformation, and he advanced the understanding and treatment of various childhood neurological conditions such as tuberous sclerosis.

Peter's greatest passions were teaching, his research and caring for the innumerable children with brain conditions who came to him for treatment. But he also enjoyed classical music, playing the flute, gardening and baking. And he took great pleasure in his three children (all of whom have pursued impressive careers) and four grandchildren. He will be remembered for his pioneering and visionary work on human synaptic development, and also by the many patients whose lives he made better and the many physicians and students whom he trained in his years at Yale and at the University of Chicago. ■

Christopher A. Walsh *is a Howard Hughes Medical Institute investigator, chief of genetics and genomics at Boston Children's Hospital and professor of paediatrics at Harvard Medical School in Boston, Massachusetts. He was a student under Peter Huttenlocher at the University of Chicago in Illinois.*
e-mail: christopher.walsh@childrens.harvard.edu

FORUM EARTH SCIENCE

The timing of climate change

An innovative assessment of climate change calculates the year in which ongoing warming will surpass the limits of historical climate variability. Three experts explain this calculation's significance compared with conventional approaches, and its relevance to Earth's biodiversity. SEE ARTICLE P.183

THE PAPER IN BRIEF

- In this issue, Mora *et al.*[1] present a novel characterization of climate change.
- The assessment is based on an estimate of when ongoing warming will continuously exceed the bounds of historical climate variability.
- Depending on the volumes of greenhouse gases emitted in the future, this could, averaged globally, occur as early as mid-century.
- The change will occur first in the tropics, for which historical variability is low and biodiversity is highest (Fig. 1).

Beyond historical variability

CHRIS HUNTINGFORD & LINA MERCADO

The debate on climate change focuses almost exclusively on mean meteorological adjustments likely to occur by the years 2020, 2050 and 2100. The first of these is an immediate timescale, which society may consider preparing for now. The year 2050 is a time by which significant change might occur, and large-scale planning to counteract this should, arguably, commence soon. It may take the next three decades to implement major adaptation schemes, or to devise mitigation solutions that could potentially include geoengineering.

Mora and colleagues' study provides two refreshing advances to our understanding of climate change. First, it concentrates on climate variability instead of overall mean change. Second, it identifies the year in which climate change will cross the threshold whereby events that are currently regarded as extreme become the norm. This may seem trivial — the authors are simply reinterpreting and inverting the same 'change versus year diagrams', but for a prescribed threshold. However, the subtle change of using scientifically based potential years of concern, as opposed to standard prescribed years such as 2050, may refocus the climate-policy debate.

Increasing emphasis is being placed on climate models in the prediction of not only average changes in weather conditions, but also properties of extreme climatic events. Risks associated with any increased frequency of such events have been well documented in both academic circles and the mainstream media. For instance, the hydrological cycle could adjust such that some regions receive more pronounced rainfall events, triggering flooding and landslides; measurements already suggest rainfall intensification in some areas[2]. Elsewhere, more frequent droughts could reduce crop yields. And how should cities be designed to deal with recurrent heatwaves? Such issues led the Intergovernmental Panel on Climate Change to commission a specific report on the current understanding of climate extremes[3].

There are also concerns about how these extremes might influence ecosystems and biodiversity. Mora and colleagues' analysis adopts an adaptation hypothesis that the bounds of ecosystem resilience might be tightly tied to historical exposure to climatic variation. If this proves to be the case, it could have surprising implications. For regions of relatively low interannual variability, such as the tropics, ecological damage could occur in response to a smaller amount of climate change than previously thought. This might be true even in the absence of an increase in extremes, because average change alone could cause persistent crossing of historical warming limits. Indeed, recent work[4] suggests that, in the past few decades, interannual temperature variation has actually decreased for the tropics. In other places, more frequent extremes may soon routinely push beyond the limits of adaptation. Some argue that more extremes are already starting to decrease the ability of terrestrial ecosystems to sequester carbon and offset carbon dioxide emissions[5].

For a 'business-as-usual' emissions scenario, averaging across all regions and climate models, Mora *et al.* estimate 2047 to be the mean year in which warming will begin to reach almost continuously beyond historical bounds, potentially affecting ecosystem health drastically. If the bounds on adaptation are defined by pre-industrial variability only, an even earlier date applies. Although continuing improvements to climate models will refine these timings, there is a more fundamental problem. At present, very few large-scale ecological models account for potential acclimation to future climatic conditions. This is a short-term adjustment — occurring over minutes to weeks — made by an organism in response to any new climatic conditions to improve its performance, thereby offsetting possible damage. The main reason for this modelling omission is a general paucity of physiological data, especially for tropical ecosystems, to parameterize such effects should they exist. Arguably, even less seems to be known about large-scale climate–biodiversity interactions and feedbacks, preventing interpretation of climate-model diagnostics in the context of biodiversity stability.

Chris Huntingford *is at the Centre for Ecology & Hydrology, Wallingford OX10 8BB, UK.* **Lina Mercado** *is at the College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4RJ, UK.*
e-mails: chg@ceh.ac.uk; l.mercado@exeter.ac.uk

Tropical diversity countdown

ERIC POST

Rather than emphasizing general projections of expected temperature increases over the next century, Mora *et al.* focus on estimating precisely when continuous deviation from mean climatic conditions that have prevailed over the past century and a half is likely to occur. Putting numbers on their 'timing of climate departure' may prove a tricky business because detractors will surely concentrate on the precise estimates of years of


RISER/MINT IMAGES/FRANS LANTING/GETTY

Figure 1 | The tropical rainforest. Tropical regions, which are characterized by low levels of climatic variability, may be the first to experience deviation beyond the bounds of long-term past climatic conditions[1].

departure so derived. But the approach provides an important complement to the estimation of climate-change velocities[6], and, when applied regionally, may have implications for understanding the role of climate change in biodiversity loss.

As all ecologists know, the increase in species diversity from the poles to the tropics is matched by increases in temperature and precipitation, and by a decline in seasonal variation in climate[7]. However, the importance of climate history and climatic gradients in the generation of diversity gradients remains obscure[8], despite improvements in model complexity[9].

Because the processes that generate and maintain diversity operate over disparate scales of space and time, we might expect the roles of variation in climatic conditions through space and time to differ between those processes as well[10]. Hence, where the maintenance of greater diversity in the tropics is concerned, the comparative stability of climatic conditions in the tropics relative to temperate and polar regions may be invoked more convincingly. This is because tropical species, especially the highly diverse group known as terrestrial ectotherms — which includes amphibians and reptiles — are characterized by narrow ranges of thermal tolerance. This indicates that these creatures have a high degree of specialization for relatively constant thermal conditions compared with organisms adapted to environmental conditions at higher latitudes[11].

Mora and colleagues' study suggests that it is these low-latitude specialists that will be the first to experience deviation of climatic conditions from long-term average conditions. Therefore, regardless of the precision of the estimated timing of climate departure itself, the tropics will reach this state first, an intriguing insight. This conclusion derives from the authors' emphasis on deviation from existing conditions as opposed to the absolute magnitude of expected warming, which is and will continue to be greatest at high northern latitudes.

Departure is expected to occur first in the tropics precisely because this region has had a relatively invariable climate over the period 1860–2005, which the authors use as a baseline. Thus, even with relatively little absolute warming, conditions in the tropics will rapidly deviate from historical baseline variability. And it is this deviation that makes sense biologically when considering the conservation implications of climate change for species with narrow physiological bounds of thermal tolerance[11]. If the assessment by Mora *et al.* proves accurate, conservation practitioners take heed — the climate-change race is not only on, it is fixed, with the extinction finish line looming closest for the tropics. ■

Eric Post *is in the Polar Center and Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA.*
e-mail: esp10@psu.edu

1. Mora, C *et al. Nature* **502,** 183–187 (2013).
2. Min, S.-K., Zhang, X., Zwiers, F. W. & Hegerl, G. C. *Nature* **470,** 378–381 (2011).
3. IPCC *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds Field, C. B. *et al.*) (Cambridge Univ. Press, 2012).
4. Huntingford, C., Jones, P. D., Livina, V. N., Lenton, T. M. & Cox, P. M. *Nature* **500,** 327–330 (2013).
5. Reichstein, M. *et al. Nature* **500,** 287–295 (2013).
6. Loarie, S. R. *et al. Nature* **462,** 1052–1055 (2009).
7. Gaston, K. J. *Nature* **405,** 220–227 (2000).
8. Pimm, S. L. & Brown, J. H. *Science* **304,** 831–833 (2004).
9. Gotelli, N. J. *et al. Ecol. Lett.* **12,** 873–886 (2009).
10. Post, E. *Ecology of Climate Change: The Importance of Biotic Interactions* (Princeton Univ. Press, 2013).
11. Tewksbury, J. J., Huey, R. B. & Deutsch, C. A. *Science* **320,** 1296–1297 (2008).

PALAEONTOLOGY

A jaw-dropping fossil fish

The ancestors of modern jawed vertebrates are commonly portrayed as fishes with a shark-like appearance. But a stunning fossil discovery from China puts a new face on the original jawed vertebrate. SEE ARTICLE P.188

MATT FRIEDMAN & MARTIN D. BRAZEAU

Jawed vertebrates, or gnathostomes, comprise the vast majority of backboned animals. Understanding their great success and the evolution of their anatomical marvels lies in reconstructing the nature of their most recent common ancestor. An entrenched assumption of classical comparative anatomy[1], which has gained some recent palaeontological support[2,3], is that this precursor resembled a shark. This hypothesis implies that the anatomical peculiarities of bony fishes — the group that would eventually give rise to ourselves — are specializations that arose later in vertebrate evolutionary history. On page 188 of this issue, Zhu *et al.*[4] present a 419-million-year-old fossil and accompanying analysis that turn this view on its head*.

The deepest split in the family tree of modern gnathostomes divides the cartilaginous fishes (chondrichthyans, which include sharks, rays and ratfishes) from the much more species-rich bony fishes (osteichthyans, comprising ray-finned fishes, lungfishes, coelacanths and terrestrial vertebrates). Among other differences, the two groups show contrasting conditions of the external skeleton: cartilaginous

*This article and the paper under discussion were published online on 25 September 2013.

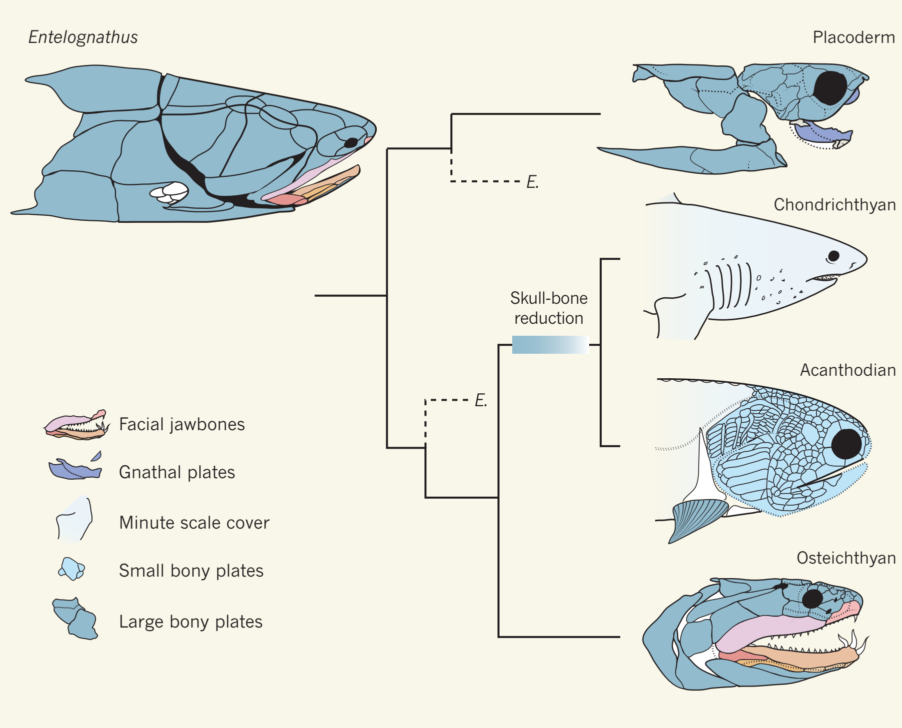


Figure 1 | Evolution of the jawed-vertebrate face. Osteichthyans are the bony fishes, from which other bony vertebrates (including amphibians, reptiles, birds and mammals) evolved. These animals have large, differentiated skull bones, whereas the skulls of the fossil acanthodians and the cartilaginous chondrichthyans (which include present-day sharks) bear numerous small bones or tiny scales. Placoderms also have large skull bones, but these extinct animals branch outside the evolutionary split between the cartilaginous and bony fishes, so the traditional view of the evolution of the modern bony-vertebrate skull and face is that it progressed from small bones to larger consolidated plates. This is challenged by Zhu and colleagues' report[4] of the fossil *Entelognathus primordialis*, which exhibits anatomical features of both osteichthyans and placoderms. The exact evolutionary position of *Entelognathus* is not certain (dotted lines), but the authors place it outside the modern jawed-vertebrate lineage. This, combined with a revised placement of the problematic acanthodians, suggests that the condition displayed by sharks and their kin is evolutionarily novel, rather than ancestral.

fishes bear a hide of tiny tooth-like scales, whereas bony fishes are clad in plate-like bones that cover the head, jaw and shoulder (Fig. 1).

These two groups diverged from a single common ancestor — neither group gave rise to the other. Consequently, it is impossible to distinguish which of these skeletal architectures is primitive on the basis of living groups alone. Despite this, there is an established tradition of considering sharks to be especially primitive, and the assumption that their anatomy approximates that of the ancestral jawed vertebrate has given rise to classic scenarios on the origins of teeth, jaws and skulls[1,5].

Palaeontological data offer a unique opportunity to test such views, and much attention has recently been lavished on an assemblage of extinct, superficially shark-like fossil fishes called the acanthodians. Analyses have interpreted[2,3] acanthodians as a hotchpotch of early bony and cartilaginous fishes, thereby lending support to the shark-like-ancestor hypothesis.

Against this backdrop, Zhu *et al.* cast *Entelognathus primordialis*, the latest discovery from a remarkable Chinese site of Silurian age (between 443 million and 419 million years old) that previously yielded the oldest known complete fossil bony fishes[6]. Articulated gnathostome remains are vanishingly rare in Silurian rocks[6], so Zhu and colleagues' find of a new, intact fish is newsworthy, if not unexpected.

At first glance, *Entelognathus* looks like a placoderm — an extinct, primitive form of armoured, jawed fish that branches outside the evolutionary split between the cartilaginous and bony fishes[2,3,7] (Fig. 1). When examined from the side, however, *Entelognathus* reveals itself as anything but expected. Absent are gnathal plates — simple jawbones characteristic of placoderms. Instead, the mouth is rimmed with bones that integrate with the cheek plates, the lower jaw is composed of an elongated 'box' of bony plates and cartilage, and the throat and gills are clad in a series of articulating plates. Both in the overall pattern and the specific detail of these plates, the fossil showcases traits that were once considered diagnostic of bony fishes[8], and entirely unknown in placoderms. *Entelognathus*, it seems, is a placoderm with a bony-fish-like grin.

So what to make of this piscine mash-up? Either the osteichthyan-like features of *Entelognathus* are antecedents of those same structures in bony fishes, or they have arisen independently in the two lineages. Zhu and colleagues' results do not unambiguously support one alternative over the other, but two things are clear from the various possibilities proposed in their evolutionary tree (see Fig. 6 of the paper[4]). First, *Entelognathus* always branches outside the radiation of living jawed vertebrates, meaning that key components of the osteichthyan face are no longer unique innovations of that group[8]. Second, acanthodians — that pivotal assortment of extinct shark-like fishes — are shifted, en masse, to the branch containing the cartilaginous fishes. This triggers a cascade of implications. If all acanthodians are early cartilaginous fishes, then their shark-like features are not generalities of jawed vertebrates, but specializations of the cartilaginous-fish branch. The most recent common ancestor of jawed vertebrates was thus probably clad in bony armour of the sort common to both placoderms and bony fishes.

This inversion of a classic scenario in vertebrate evolution raises an obvious question: how did we get it so wrong? The status of sharks as surrogate ancestors seems well established, but this is an illusion of dogmatic repetition combined with spurious portrayals of present-day cartilaginous fishes as unchanged 'living fossils'. The popular model of a shark-like ancestor is, in the end, more a hangover of the 'great chain of being' of ancient philosophy and pre-Darwinian archetypes than a product of modern comparative biology and phylogenetic 'tree thinking'. Added to this conceptual inertia is a historically compartmentalized approach to studying early vertebrate groups[9] that made it too easy to dismiss shared similarities — the head and shoulder exoskeleton of placoderms and bony fishes, for example — as independent innovations without adequate evidence.

Over the past decade or so, new fossils[10,11] and re-examinations of old ones[2,3] have forced palaeontologists to look beyond the confines of traditional classifications and reconsider the coherence of textbook assemblages such as placoderms and acanthodians, and their relationships to extant gnathostomes[2–4,7]. Perhaps more than any of these discoveries, *Entelognathus* demands a major rethink of where fossils fit relative to modern lineages, and how these living groups came to acquire their characteristic traits. It will take time to fully digest the implications of such a remarkable fossil, but it is clear that a major reframing of our understanding of early gnathostome evolution is now in full swing. ■

Matt Friedman *is in the Department of Earth Sciences, University of Oxford, Oxford OX1 3AN, UK.* **Martin D. Brazeau** *is in the Department of Geology, Naturalis Biodiversity*

Center, 2333 CR Leiden, the Netherlands, and in the Department of Life Sciences, Imperial College London, UK.
e-mails: mattf@earth.ox.ac.uk; martin.brazeau@gmail.com

1. Goodrich, E. S. *Studies on the Structure and Development of Vertebrates* (Macmillan, 1930).
2. Brazeau, M. D. *Nature* **457,** 305–308 (2009).
3. Davis, S. P., Finarelli, J. A. & Coates, M. I. *Nature* **486,** 247–250 (2012).
4. Zhu, M. *et al. Nature* **502,** 188–193 (2013).
5. Reif, W. E. *Evol. Biol.* **15,** 278–386 (1982).
6. Zhu, M. *et al. Nature* **458,** 469–474 (2009).
7. Friedman, M. *J. Syst. Palaeontol.* **5,** 289–343 (2007).
8. Friedman, M. & Brazeau, M. D. *J. Vert. Paleontol.* **30,** 36–56 (2010).
9. Janvier, P. *Early Vertebrates* (Clarendon, 1996).
10. Zhu, M., Yu, X. & Janvier, P. *Nature* **397,** 607–610 (1999).
11. Miller, R. F., Cloutier, R. & Turner, S. *Nature* **425,** 501–504 (2003).

QUANTUM PHYSICS

Watching the wavefunction collapse

The continuous random path of a superconducting system's quantum state has been tracked as the state changes during measurement. The results open the possibility of steering quantum systems into a desired state. SEE LETTER P.211

ANDREW N. JORDAN

Continuous quantum measurements of a system gently interrogate the state of the system. The usually abrupt collapse of the wavefunction, which generally occurs upon measurement, happens gradually over a period of time. Advances in superconducting quantum systems now permit scientists to track a single 'quantum trajectory' describing the continuous random path a quantum state takes as it collapses from a quantum superposition into one of its classically permitted states. Writing in this issue (page 211), Murch *et al.*[1] report an experiment that both confirms the theory of quantum trajectories and develops a way to quantum control in solid-state systems.

The past decade has seen tremendous technological development in the area of superconducting quantum systems. Research on quantum measurements and control of single systems has until recently been dominated by the field of quantum optics, which deals with photons and atoms[2]. However, techniques to fabricate man-made atoms in the form of superconducting systems — which display discrete energy levels and the long quantum coherence times that are critical for making many operations on the systems before they become effectively classical objects — have steadily produced longer coherence times. In 2000, the typical coherence times of such superconducting quantum systems were 10 nanoseconds, whereas today the times are in excess of 100 microseconds, a 10,000-fold increase[3].

Murch and colleagues' experiment uses a relatively new kind of system called a three-dimensional transmon[3]. This system is a superconducting element that is described by a ground (lowest-energy) state and a first excited state, and is kept at a low temperature. The quantum state of the system can be detected by putting the superconducting element inside a microwave box. The detector works in a similar way to the microwave ovens sitting in many kitchens. Microwaves are sent into the box, and interact with the superconducting system. A kitchen microwave causes water molecules to quickly rotate and bump into their neighbours, so that the microwaves are absorbed and energy is directly transferred to heat the water in food.

By contrast, the superconducting transmon is fixed in place, and the microwaves are far off being in resonance with the frequency needed to drive the system between its ground and excited states, so no absorption occurs. This causes an indirect interaction between the microwave photons and the quantum system. The microwaves come back out of the box through a separate port, and the information about the quantum system is read out from a shift in phase (where a wave's peaks and troughs lie) that the waves undergo in the box. This is possible because the physics of the interaction between the microwaves and the superconducting system is well understood, and can be used to calibrate the output to tell which state the system is in.

An important element of this experiment is

Figure 1 | Of butterflies and quantum trajectories. Murch and colleagues[1] performed measurements on a quantum system that enabled them to track the continuous random path, or quantum trajectory, of the system's state as it collapsed to a final state during the measurement process. Here, measurements starting from the same initial state will realize a different quantum trajectory, eventually ending in the quantum system's ground or excited state. The process is analogous to watching butterflies that make their way, one by one, across a field from a cage to one of two nearby trees (the ground and excited states in this analogy). Each butterfly's flight pattern is like a single run of the experiment.

that it takes time to fully measure the system. Although there is information about the quantum system in the outcoming microwaves, it is only a small amount of information. This is because the microwave photons exhibit quantum noise that masks the system contribution. Such a situation is sometimes called a weakly responding detector[4] and the underlying measurements are called weak measurements[5]. Consequently, only after collecting a sufficient amount of data can one distinguish the system signal from the noise and reliably determine the system's final collapsed state.

During the process of data collection, the authors use the quantum-trajectories formalism to infer the state of the quantum system from the continuous stream of microwaves. This formalism recognizes that the process of collecting data is continuous, and therefore uses the information contained in the weak measurement to update the quantum state in light of the new information received. The fact that there is very little knowledge gained about the system in the data means that the disturbance to the system is small, so the updated state is close to the original state. This process is then repeated as more data are collected, which generates the quantum trajectory, where the state follows a diffusive motion in time. Thus, Murch and colleagues literally watch the wavefunction continuously collapse to a final state in a single experimental run.

The method of following the quantum state in time is similar to chasing butterflies released from a cage (Fig. 1). Suppose each butterfly flies in a random zig-zag pattern, but eventually makes its way across a field to land on one of two trees. If another butterfly is released with the same initial conditions, it follows a different zig-zag path. Consequently, it is not possible to predict the flight pattern of the butterfly. However, after watching enough butterflies make the journey, one can make statistical predictions about questions such as: which tree will it land on? How long will the journey take? What will the average path be, given it lands on the first tree? In the same way, although an individual quantum trajectory is a random process, similar statistical questions can be asked, such as: what is the probability that the system will end up in the excited state? What is the average time it takes to get there? What will the average quantum trajectory be, given that it ends in the excited state[6]? One of the theoretical predictions successfully checked in this work is that the quantum trajectory at a given time depends only on the total integrated detector signal up to that point in time.

When a butterfly is chased, it will alter its path in response to being pursued, opening the possibility of steering the butterfly to where you want it to go. Similarly, the fact that quantum trajectories can be monitored suggests the possibility of steering them through the use of feedback control — in which control parameters of the quantum system are dynamically changed in response to the measurement outcome. Indeed, this same research group has already published the results of an experiment implementing such control to stabilize the dynamics of a continuously measured system[7].

The published data also show the effect of measurement uncollapse in continuous measurements[8,9]. Because the detector gives only partial information about the quantum state, there is the possibility that the detector output 'reverses its judgement' about the state of the system. The authors noticed that there were several times when the quantum trajectory returned the state back to where it started. The detector measured the system for some time, partially collapsing the system's state, and then measured again for some time, uncollapsing it.

This experiment further cements solid-state systems as a central player in the manipulation and control of single quantum systems, the area of the 2012 Nobel Prize in Physics, awarded to Serge Haroche and David Wineland[10]. It opens the way to further investigations in quantum feedback and control, and it exhibits the fundamental importance of new ideas in weak and continuous quantum measurement, a field that has until recently been largely ignored in solid-state physics, but now seems to have an upward 'trajectory'. ■

Andrew N. Jordan *is in the Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627, USA, and at the Institute for Quantum Studies, Chapman University, Orange, California 92866, USA.*
e-mail: jordan@pas.rochester.edu

1. Murch, K. W., Weber, S. J., Macklin, C. & Siddiqi, I. *Nature* **502,** 211–214 (2013).
2. Wiseman, H. M. & Milburn, G. J. *Quantum Measurement and Control* (Cambridge Univ. Press, 2010).
3. Devoret, M. H. & Schoelkopf, R. J. *Science* **339,** 1169–1174 (2013).
4. Korotkov, A. N. *Phys. Rev. B* **60,** 5737–5742 (1999).
5. Aharonov, Y., Albert, D. Z. & Vaidman, L. *Phys. Rev. Lett.* **60,** 1351–1354 (1988).
6. Chantasri, A., Dressel, J. & Jordan, A. N. Preprint at http://arxiv.org/abs/1305.5201 (2013).
7. Vijay, R. *et al. Nature* **490,** 77–80 (2012).
8. Korotkov, A. N. & Jordan, A. N. *Phys. Rev. Lett.* **97,** 166805 (2006).
9. Katz, N. *et al. Phys. Rev. Lett.* **101,** 200401 (2008).
10. Georgescu, I. *Nature Phys.* **8,** 777 (2012).

MICROBIOLOGY

RNAs at fever pitch

Some normally innocuous bacteria can turn into serious pathogens. It seems that one such species, *Neisseria meningitidis*, uses three RNA-based thermosensors to escape the immune response of its human host. SEE LETTER P.237

FRANZ NARBERHAUS

The flow of genetic information, from DNA to RNA to protein, is strictly controlled, such that only a subset of genes is actively expressed under any given condition. Contact between a bacterium and a mammalian host elicits profound changes in bacterial gene expression to adjust to the new conditions and to escape detection by the host's immune system. But what host cellular parameters are monitored, and at what level does gene regulation occur? On page 237 of this issue, Loh *et al.*[1] report that *Neisseria meningitidis* registers increasing temperature as an alarm signal that indicates the risk of an inflammation-activated immune system*. They show that the bacterium uses three different temperature-sensing RNA molecules to coordinate processes involved in evading this immune response.

Neisseria meningitidis (often referred to as meningococcus) is the leading cause of septicaemia and meningitis in young children and adolescents. Paradoxically, it is a harmless commensal (resident) bacterium found in the nasopharynx of many individuals. The triggers that can lead this microorganism to enter the bloodstream and cross the blood–brain barrier, and the mechanisms by which it achieves this while escaping immune detection, are only partially understood.

To identify factors involved in these processes, Loh and collaborators screened for mutant *N. meningitidis* strains that were able to survive repeated exposure to human serum, which contains immune-defence molecules. Five out of the six resistant strains that they identified were characterized by the loss of an identical sequence of eight nucleotide bases lying in front of a genetic region involved in synthesizing the capsule — a protective polysaccharide layer that surrounds the bacterial cell.

How can such a short genetic deletion prevent killing by the immune system? The authors found that all five strains produce elevated levels of the capsule-synthesis enzyme CssA. Assuming that the deletion would destroy the recognition site for a transcription factor, they tested 38 such proteins for involvement in regulating this genetic region. None of them was involved.

*This article and the paper under discussion were published online on 25 September 2013.

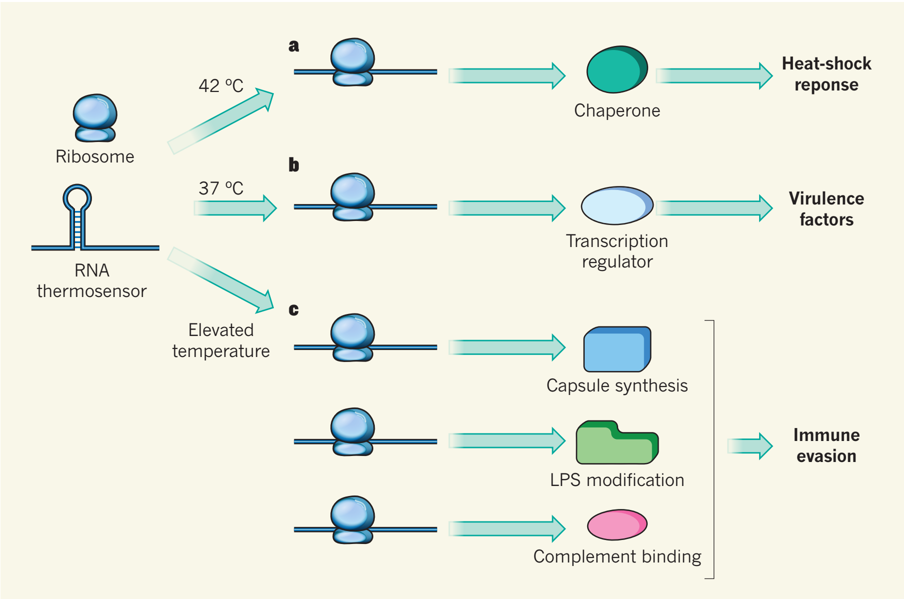


Figure 1 | RNA thermometers team up. Several bacterial RNA molecules have been identified that contain a temperature-responsive region referred to as an RNA thermosensor. Formation of a hairpin structure at this region prevents binding by the ribosome, which performs protein translation. But this structure gradually melts with increasing temperature, exposing the ribosome binding site and allowing the RNA to be translated. **a**, Several RNA thermosensors are part of RNAs encoding chaperone proteins that function in the heat-shock response (at temperatures of 42 °C or higher). **b**, Other such sensors are found in transcripts encoding proteins that regulate the expression of virulence-factor genes in response to host body temperature. **c**, Loh *et al.*[1] find that three separate RNA thermosensors function in the bacterium *Neisseria meningitidis* to control capsule synthesis, modification of membrane lipopolysaccharide (LPS) and recruitment of the complement-binding protein factor H. These sensors respond to elevated host temperature to achieve bacterial evasion of the human immune response.

This strongly suggested that the regulation occurs after the RNA has been transcribed, which led the authors to identify a surprisingly simple effect of the 8-nucleotide deletion. They show that the normal RNA sequence forms a stable hairpin structure that prevents binding of the ribosome (the molecular machine that enacts protein translation). But the shorter RNA structure in the resistant strains is more labile, permitting ribosome access and leading to enhanced capsule synthesis.

RNA structures that are similarly locked in an inert conformation are frequently found upstream of temperature-regulated genes. These 'RNA thermosensors' adopt structures that prevent translation at low temperatures, but increasing temperatures cause their structure to melt, liberating the ribosome-binding site and permitting protein synthesis[2]. Several of these molecules are chaperone RNAs that provide an immediate response to life-threatening heat-shock conditions[3,4] (Fig. 1a). Others, such as the RNA thermosensors in the bacteria *Listeria monocytogenes* and *Yersinia pseudotuberculosis*, respond to host body temperature by controlling the production of a crucial regulator of virulence genes[5,6] (Fig. 1b). As expected, the *N. meningitidis* RNA structure discovered by Loh *et al.* turned out to be temperature sensitive: production of the capsule-synthesis enzyme gradually increased with increasing temperature, consistent with the zipper-like melting that has been demonstrated for other RNA thermosensors[7,8].

In addition to the capsule, other factors help meningococcus to avoid immune killing[9]. Fascinatingly, Loh and co-workers reveal that RNA thermosensing also regulates the synthesis of two of these factors: factor H binding protein, which regulates the complement pathway of the immune system, and Lst, which modifies the lipopolysaccharide molecules in the outer membrane of the bacterium. Thus, in an unprecedented scenario, three independent RNA thermosensors join forces to counteract immune killing in this important human pathogen (Fig. 1c).

It is puzzling that the nucleotide sequences and predicted inhibitory structures of all three thermosensors are so different, because this suggests that they have evolved independently. Detailed structural studies will be needed to confirm whether the computer-predicted structures presented by Loh *et al.* are correct, and to determine exactly how the RNA molecules melt. It will also be interesting to learn whether the three thermosensors respond simultaneously or in a sequential fashion.

So why has a commensal bacterium living in the human throat established such exquisite temperature surveillance mechanisms? A rise in the human body temperature is typically associated with inflammation, and in the meningococcal habitat this can be provoked by an infection with influenza virus or other respiratory pathogens. The concomitant immune response could eradicate the harmless bystander *N. meningitidis*, and so inducing immune-evasion mechanisms as soon as the host temperature rises seems like a logical survival strategy. Moreover, the bacterium would benefit from its temperature-induced protective capsule once it gains access to the bloodstream.

Should it not then be advantageous to be permanently armed? Apparently not, according to Loh and colleagues' finding that most strains of *N. meningitidis* isolated from clinical samples, including several highly virulent strains, use temperature-regulated capsule synthesis. This is consistent with the idea that the production of virulence factors comes with a fitness cost[10] and that protective measures are taken only when the life of the bacterium is at stake.

The discovery of three RNA thermosensors raises the question of how prevalent such sensors are in nature. By far the most common genes known to be controlled by RNA thermosensors are those that encode heat-shock and virulence proteins[2]. Recently, however, an adenine-sensing riboswitch was found to be markedly temperature responsive[11]. Riboswitches are metabolite-binding RNA elements that control numerous metabolic pathways[12]. They are commonly believed to exist in two alternative conformations, either with or without their ligand. The adenine-binding riboswitch of the human pathogen *Vibrio vulfnificus*, however, adopts three distinct states[11], and the ambient temperature shifts the equilibrium between two ligand-empty states that have either low or high ligand affinity. This finding, together with Loh and colleagues' report, suggests that further temperature-sensitive RNAs might show up in unexpected places. ■

Franz Narberhaus *is at the Institute of Microbial Biology, Ruhr University Bochum, Bochum 44780, Germany.*
e-mail: franz.narberhaus@rub.de

1. Loh, E. *et al. Nature* **502,** 237–240 (2013).
2. Kortmann, J. & Narberhaus, F. *Nature Rev. Microbiol.* **10,** 255–265 (2012).
3. Waldminghaus, T., Fippinger, A., Alfsmann, J. & Narberhaus, F. *Biol. Chem.* **386,** 1279–1286 (2005).
4. Kortmann, J., Sczodrok, S., Rinnenthal, J., Schwalbe, H. & Narberhaus, F. *Nucleic Acids Res.* **39,** 2855–2868 (2011).
5. Johansson, J. *et al. Cell* **110,** 551–561 (2002).
6. Böhme, K. *et al. PLoS Pathog.* **8,** e1002518 (2012).
7. Chowdhury, S., Maris, C., Allain, F. H.-T. & Narberhaus, F. *EMBO J.* **25,** 2487–2497 (2006).
8. Rinnenthal, J., Klinkert, B., Narberhaus, F. & Schwalbe, H. *Nucleic Acids Res.* **38,** 3834–3847 (2010).
9. Schneider, M. C., Exley, R. M., Ram, S., Sim, R. B. & Tang, C. M. *Trends Microbiol.* **15,** 233–240 (2007).
10. Sturm, A. *et al. PLoS Pathog.* **7,** e1002143 (2011).
11. Reining, A. *et al. Nature* **499,** 355–359 (2013).
12. Serganov, A. & Nudler, E. *Cell* **152,** 17–24 (2013).

NATURE

50 Years Ago

'Evidence for life on Mars' — My opinion is that the evidence has been distorted and that a re-examination is required. ... As an alternative to the interpretation of life on Mars, I prefer an inorganic explanation based on ideas of McLaughlin and Kuiper. Assuming volcanoes, both active and dead, and suitable winds, the phenomena can be interpreted in a reasonably consistent manner. ... The question of life on Mars has intrigued scientists for decades. Owing to the possibility of being able to verify directly the life hypothesis in the very near future it is becoming a subject of widespread study. ... On the basis of this brief discussion it is suggested that the life interpretation is very tenuous and that inorganic interpretations be given greater credence.
From *Nature* 12 October 1963

100 Years Ago

The National Gas Exhibition at Shepherd's Bush, which will be open during the whole of October, affords the best object lesson in gas lighting that the public has ever had the opportunity of studying ... In domestic heating the grasping of the conditions necessary to make gas a hygienic domestic fuel has been the great factor that has led to progress. In the early days of the gas fire, only 25 to 33 per cent. of the heat was given out as radiant heat, and convection was relied upon largely to give the heating effect, this giving hot air to breathe, and at the same time leaving the objects in the room often so far below the body temperature as to lead to chill ... Another very suggestive exhibit is a series of compartments illustrating the effect of the colour and surface of wall-papers on the amount of illumination obtained from equal sources of light.
From *Nature* 9 October 1913

CELL BIOLOGY

Molecular clearance at the cell's antenna

The degradative process known as autophagy is a cellular quality-control mechanism that is associated with many clinical disorders. It emerges that autophagy and the cell's primary cilium regulate each other. SEE ARTICLE P.194 AND LETTER P.254

VALENTINA CIANFANELLI & FRANCESCO CECCONI

Long considered a vestigial organelle, the primary cilium has lately been identified as the cell's sensory antenna. It extends from the cell surface to the extracellular space, and so is ideally positioned to receive signals from outside. On sensing such signals, it integrates and transduces them to coordinate an appropriate cellular response and decide the cell's fate[1]. The fact that the primary cilium is present in most human cells, together with the discovery of its sensory function and the identification of cilium-related human disorders that include polycystic kidney disease, retinal degeneration and hydrocephalus, have all contributed to the recent explosion of interest in this organelle. In this context, the findings of Pampliega *et al.*[2] and Tang *et al.*[3], published in this issue, are of great importance*. They describe a role for autophagy — a pathway for the degradation of cellular components[4] — in the development of the primary cilium (ciliogenesis) and, reciprocally, the involvement of this organelle in regulating autophagy.

The primary cilium consists of a backbone of microtubules arranged into a structure called the axoneme, which is covered by the cell membrane. It grows from another organelle known as the basal body (Fig. 1). During ciliogenesis, intracellular vesicles recruit protein building blocks and shuttle them to the growing cilium[5].

*This article and the papers under discussion were published online on 2 October 2013.

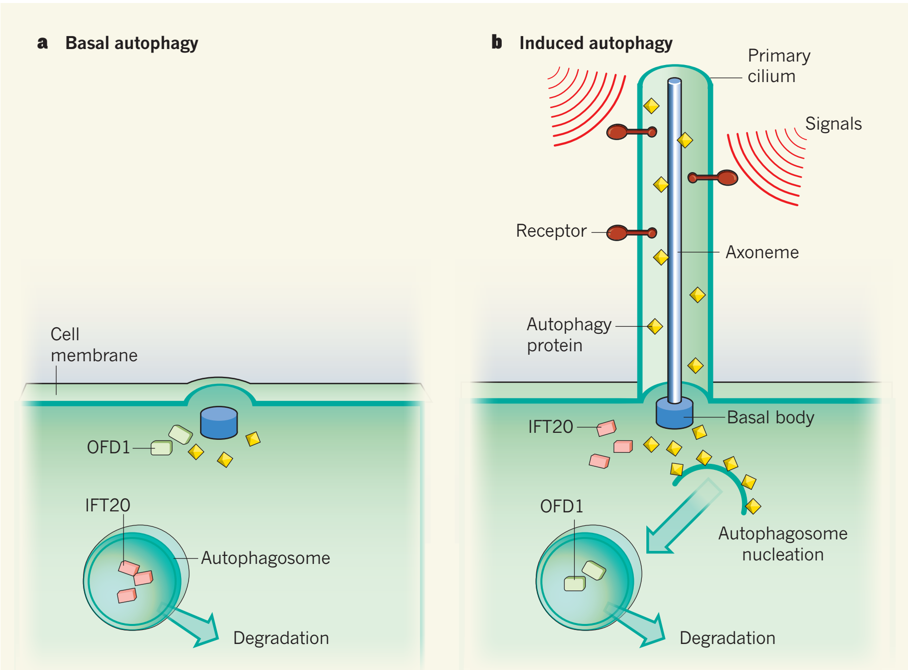


Figure 1 | Ciliogenesis and autophagy. Two studies[2,3] find that autophagy participates in ciliogenesis, and that the primary cilium is in turn necessary for full induction of autophagy. **a**, Specifically, in basal autophagy, ciliogenesis is prevented by the sequestering of IFT20 (a protein that is essential for cilium formation) in autophagosomes and by its autophagic degradation. OFD1, a protein that inhibits cilium formation, is not affected by basal autophagy. **b**, By contrast, autophagy induced by serum deprivation promotes ciliogenesis through inactivation of OFD1, while sparing IFT20. The primary cilium consists of the axoneme covered by the cell membrane and is attached to the basal body. Autophagy proteins also localize at this sensory organelle, probably contributing to cilium organization. At the basal body, signals from the cilium stimulate autophagosome nucleation.

Pampliega *et al.* and Tang and colleagues identify two targets of autophagy among the proteins that participate in ciliogenesis. When cells are deprived of serum (a stimulus that promotes both autophagy and ciliogenesis)[4,5], 'induced' autophagy triggers ciliogenesis by degrading OFD1 — a protein that inhibits cilium formation[3]. By contrast, basal autophagy, which occurs under normal nutrient conditions, prevents ciliogenesis by sparing OFD1 (ref. 3) and digesting IFT20 (ref. 2), a protein essential for both cilium formation and length regulation[5] (Fig. 1). These findings are complementary in that they describe a regulation of ciliogenesis by two different phases of autophagy. This 'double lock' hints at the importance of proper cilium regulation for cell survival.

The autophagy machinery seems to be strategically present at the cilium. Both research teams show that components of this machinery co-localize with its target substrates (OFD1 and IFT20) and that this co-localization is crucial for regulating ciliogenesis. The mechanisms underlying the selective degradation of ciliary proteins by basal and induced autophagy remain elusive, but could involve post-translational modification of the proteins or protein–protein interactions. Because elucidation of such mechanisms could contribute to the design of therapies for cilium-related diseases[5], further studies of the post-translational modification (such as ubiquitination or phosphorylation[6]) of ciliary proteins that are autophagy substrates, and of interactions between these proteins and cargo-recognition proteins of the autophagy pathway (such as p62 and NBR1 proteins[7]), could yield useful results.

Notably, Pampliega *et al.* also propose that ciliogenesis regulates induced autophagy. This could at least partly involve the cilium membrane acting as a nucleation site for the formation of pre-autophagosomes and subsequently autophagosomes (membrane-bounded vesicles that carry cellular components destined for degradation). In support of this hypothesis, the authors find that proteins participating in autophagosome nucleation reside near the cilium membrane, associating with the axoneme[2], or at the base of the cilium[2,3]. This theory is also consistent with the discovery that the cell membrane is a source of autophagosomes[8]. However, beclin 1, a protein necessary for autophagosome nucleation, is missing at the cilium. This is puzzling, and suggests the existence of alternative protein complexes for autophagosome nucleation in this specialized cellular compartment.

Nonetheless, other possible functions of the autophagy components at the primary cilium should be investigated. For example, the fact that vesicles bearing autophagy proteins are found at the cilium could mean that autophagy mediates the internalization of receptors located at the cilium membrane. Indeed, a role for autophagy in mediating such internalization has been documented in epithelial cells[9].

Another intriguing possibility is that autophagosomes are responsible for delivering proteins to the cilium membrane or for releasing their cargo in the extracellular environment, as described in other contexts[10]. Such roles for autophagy would imply that this degradative process directly impinges on the signals captured or delivered by the cilium. Finally, autophagy-independent functions of autophagy proteins at the axoneme cannot be ruled out; for example, they might have a structural role. Some of these proteins carry a cilium-targeting sequence[2], and experimental modification of this sequence to prevent the proteins from localizing to the axoneme could aid analysis of their function at the cilium.

The primary cilium is a signalling platform that senses different stimuli in different cell types. Its involvement in the regulation of autophagy opens up the possibility that atypical stimuli that are transduced through this organelle, such as changes in temperature, light, pressure and osmolarity, also affect autophagy. The signalling pathways that act on autophagy downstream of the primary cilium could be cell-type specific, and studies in diverse cell types could confirm this. For now, Pampliega and colleagues show that the Hedgehog signalling pathway mediates ciliary regulation of autophagy in mouse embryonic fibroblasts.

The results obtained by Pampliega *et al.* and Tang *et al.* offer fresh prospects for cilium biology, and their data may have far-reaching implications for unravelling the mechanisms behind cilium- and autophagy-related disorders. Moreover, modulating local autophagy with drugs — a strategy already pursued in fighting cancer and neurodegeneration[11] — may turn out to be a useful approach for switching our cellular antennas on or off on demand. ■

Valentina Cianfanelli *and* **Francesco Cecconi** *are in the Department of Biology, University of Tor Vergata, Rome 00133, Italy, and in the Department of Neuroscience, IRCCS Santa Lucia Foundation, Rome.* **F.C.** *is also at the Unit of Cell Stress and Survival, Danish Cancer Society Research Center, Copenhagen, Denmark.*
e-mail: francesco.cecconi@uniroma2.it

1. Satir, P., Pedersen, L. B. & Christensen, S. T. *J. Cell Sci.* **123,** 499–503 (2010).
2. Pampliega, O. *et al. Nature* **502,** 194–200 (2013).
3. Tang, Z. *et al.* **502,** 254–257 (2013).
4. Parzych, K. R. & Klionsky, D. J. *Antioxid. Redox Signal.* http://dx.doi.org/10.1089/ars.2013.5371 (2013).
5. Ishikawa, H. & Marshall, W. F. *Nature Rev. Mol. Cell Biol.* **12,** 222–234 (2011).
6. McEwan, D. G. & Dikic, I. *Trends Cell Biol.* **21,** 195–201 (2011).
7. Lamark, T., Kirkin, V., Dikic, I. & Johansen, T. *Cell Cycle* **8,** 1986–1990 (2009).
8. Cuervo, A. M. *Nature Cell Biol.* **12,** 735–737 (2010).
9. Thoresen, S. B., Pedersen, N. M., Liestøl, K. & Stenmark, H. *Exp. Cell Res.* **16,** 3368–3378 (2010).
10. Dupont, N. *et al. EMBO J.* **30,** 4701–4711 (2011).
11. Choi, A. M. K., Ryter, S. W. & Levine B. *N. Engl. J. Med.* **368,** 651–662 (2013).

METABOLIC SYNDROME

F stands for fructose and fat

Fructose and glucose have the same caloric value, but the two sugars are metabolized differently. It emerges that mice that cannot metabolize fructose are healthier when placed on carbohydrate-rich diets.

COSTAS A. LYSSIOTIS & LEWIS C. CANTLEY

A drastic increase in dietary sugar consumption in the western world during the past four decades has been paralleled by epidemics of obesity and metabolic syndrome, suggesting a cause-and-effect relationship. Yet the relative contribution of individual sugars — as opposed to total caloric intake — to this epidemic remains controversial. For instance, increased intake of fructose, which is enriched in soft drinks and processed foods, has been proposed[1] to greatly contribute to these disorders. However, this proposal has not been universally embraced[2]. Two studies by Johnson and colleagues, published in *Hepatology* (Ishimoto *et al.*[3]) and *Nature Communications* (Lanaspa *et al.*[4]), now investigate the role of fructose metabolism in obesity and metabolic syndrome using a mouse model that cannot metabolize this sugar. The results strongly support the theory that fructose overconsumption has toxic effects.

Dietary sugar encompasses several carbohydrates. Most often, however, it describes starch, sucrose and high-fructose corn syrup, each of which is composed of glucose with or without fructose: starch, found in bread and rice, is a glucose polymer; sucrose (table sugar) is a disaccharide made up of glucose

and fructose; and high-fructose corn syrup, a common constituent of soft drinks, is a mixture of approximately 40% glucose and 60% fructose[5]. From an energetic standpoint, a molecule of glucose has the same caloric value as a molecule of fructose. However, the human body treats these carbohydrates quite differently, raising questions about their individual roles in obesity and metabolic syndrome[1].

In general, glucose is used directly by tissues such as the muscles and brain as an energy source. Excess glucose is stored in the liver as glycogen (a glucose polymer) but can also be converted into fructose by the polyol biochemical pathway (Fig. 1). By contrast, fructose is almost exclusively metabolized by the liver. In this organ, ketohexokinase (KHK) — a liver-specific fructose-metabolizing enzyme also known as fructokinase — traps fructose in liver cells as fructose 1-phosphate. Unlike fructose 6-phosphate (an isomer of fructose 1-phosphate that participates in the biochemical pathway of glycolysis), fructose 1-phosphate can bypass a major regulatory step in glycolysis that generates fructose 1,6-bisphosphate through the action of the energy-sensitive enzyme phosphofructokinase. Thus, fructose can be converted into fat, unfettered by the cellular controls that prevent unrestrained lipid synthesis from glucose[1,6].

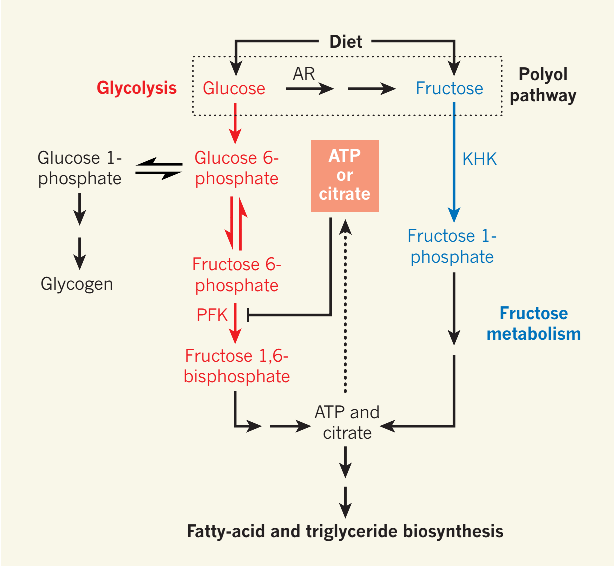


Figure 1 | Sugar metabolism. Glucose and fructose are dietary sugars. Glucose is mainly metabolized by glycolysis and can be regulated through feedback inhibition by ATP or citrate, which redirects glucose towards storage as glycogen. In liver cells expressing the enzyme ketohexokinase (KHK), fructose is metabolized to fatty acids, triglyceride head groups (glycerol, not shown) and triglycerides — a biosynthetic pathway that is not regulated by the ATP- or citrate-feedback mechanisms. Glucose can also be converted into fructose through the polyol pathway, which involves the action of the enzyme aldose reductase (AR). In the absence of feedback inhibition, glucose metabolites can also be used to generate fatty acids and triglycerides. PFK, phosphofructokinase.

By this logic, diets high in fructose could cause excess fat accumulation in the liver, leading to the liver disorders fatty liver disease, steatohepatitis and, ultimately, cirrhosis. Liver fat could also be released into the circulation and taken up by fat cells in other tissues, resulting in obesity. Furthermore, the circulating fat could accelerate the onset of cardiovascular disease, insulin resistance and type 2 diabetes. So fructose over-consumption may be at the heart of metabolic syndrome, which has also been linked to poor outcome of a wide range of cancers[7].

Indeed, there are epidemiological links between fructose consumption, obesity and metabolic syndrome[8]. Moreover, several studies[6,9,10] have demonstrated that excess fructose consumption causes features of metabolic syndrome in laboratory animals and humans. For example, in overweight humans, a diet high in fructose (25% of total caloric intake) promoted insulin resistance, elevated levels of the lipid triglyceride and visceral obesity — features not observed in overweight individuals on an otherwise identical glucose-based diet[6]. Such studies strongly suggest that over-eating fructose (as opposed to glucose) promotes metabolic syndrome. Yet, despite this growing body of circumstantial evidence, and the logic behind the argument, the relative contribution of fructose to this condition has remained unproven.

Ishimoto and colleagues set out to directly assess the role of fructose metabolism on features of metabolic syndrome, using mice lacking KHK and, as such, incapable of processing fructose. Their findings are consistent with those of several previous studies: wild-type mice fed a western diet (one that is high in fat and fructose, in which the fructose comes from sucrose) developed severe fatty liver and liver inflammation (disorders collectively known as non-alcoholic steatohepatitis) along with liver fibrosis. By contrast, KHK-deficient mice fed the same diet were protected from liver inflammation and fibrosis, and developed only mild fatty liver. Because the authors demonstrate that the mutant and wild-type mice had equivalent caloric intake, the protection afforded by the mutant animals' inability to process fructose is direct evidence that this sugar has a role in exacerbating specific features of metabolic syndrome.

Lanaspa *et al.* take the matter a step further, examining the effect of carbohydrate-rich diets devoid of fructose in KHK-deficient mice. Intriguingly, these animals were also protected from the adverse effects of excess glucose consumption. This makes sense, given that in the liver the polyol pathway converts excess glucose into fructose, which is stored as fat only in the presence of KHK (Fig. 1). In fact, the authors validate the dependence of fructose synthesis on this pathway in mice lacking the polyol-pathway enzyme aldose reductase; these animals were also protected from glucose-induced fatty liver.

Whether fructose biosynthesis from a high-glucose diet is also relevant to metabolic syndrome in humans remains unknown. Regardless of the answer, the current papers make a strong case for the toxic effects of excess carbohydrate intake, placing fructose metabolism in the crosshairs. From an evolutionary perspective, enhanced fructose conversion to fat may have been advantageous. For instance, because fruits ripen at the end of the growth season, starch conversion into fructose and, in turn, storage of the abundant fructose as fat rather than glycogen could have facilitated survival through the ensuing months of low food availability.

In the western world today, however, food is rarely limiting at any time of the year, so the fat stored from the consumption of fructose remains unused. Moreover, many of the foods that are available at low cost year-round are those with the highest fructose content (sugary drinks and processed foods). That only about half of the calories in sucrose and in high-fructose corn syrup are immediately available as glucose for use by the muscles and brain may also contribute to our tendency to over-eat these sugars in order to maintain blood glucose levels. Thus, despite being calorically equivalent, a basic biochemical understanding of sugar metabolism clearly illustrates that fructose and glucose are not equal. ■

Costas A. Lyssiotis *and* **Lewis C. Cantley** *are in the Department of Medicine, Weill Cornell Medical College, New York, New York 10065, USA.*
e-mails: col2007@med.cornell.edu; lcantley@med.cornell.edu

1. Lustig, R. H. *J. Am. Diet. Assoc.* **110,** 1307–1321 (2010).
2. Feinman, R. D. & Fine, E. J. *Nutr. Metab.* **10,** 45 (2013).
3. Ishimoto, T. *et al. Hepatology* http://dx.doi.org/10.1002/hep.26594 (2013).
4. Lanaspa, M. A. *et al. Nature Commun.* **4,** 2434 (2013).
5. Ventura, E. E., Davis, J. N. & Goran, M. I. *Obesity* **19,** 868–874 (2011).
6. Stanhope, K. L. *et al. J. Clin. Invest.* **119,** 1322–1334 (2009).
7. Khandekar, M. J., Cohen, P. & Spiegelman, B. M. *Nature Rev. Cancer* **11,** 886–895 (2011).
8. Dhingra, R. *et al. Circulation* **116,** 480–488 (2007).
9. Teff, K. L. *et al. J. Clin. Endocrinol. Metab.* **94,** 1562–1569 (2009).
10. Jürgens, H. *et al. Obes. Res.* **13,** 1146–1156 (2005).

L.C.C. declares competing financial interests. See go.nature.com/3dj3rg for details.

doi:10.1038/nature12540

The projected timing of climate departure from recent variability

Camilo Mora[1], Abby G. Frazier[1], Ryan J. Longman[1], Rachel S. Dacks[2], Maya M. Walton[2,3], Eric J. Tong[3,4], Joseph J. Sanchez[1], Lauren R. Kaiser[1], Yuko O. Stender[1,3], James M. Anderson[2,3], Christine M. Ambrosino[2,3], Iria Fernandez-Silva[3,5], Louise M. Giuseffi[1] & Thomas W. Giambelluca[1]

Ecological and societal disruptions by modern climate change are critically determined by the time frame over which climates shift beyond historical analogues. Here we present a new index of the year when the projected mean climate of a given location moves to a state continuously outside the bounds of historical variability under alternative greenhouse gas emissions scenarios. Using 1860 to 2005 as the historical period, this index has a global mean of 2069 (±18 years s.d.) for near-surface air temperature under an emissions stabilization scenario and 2047 (±14 years s.d.) under a 'business-as-usual' scenario. Unprecedented climates will occur earliest in the tropics and among low-income countries, highlighting the vulnerability of global biodiversity and the limited governmental capacity to respond to the impacts of climate change. Our findings shed light on the urgency of mitigating greenhouse gas emissions if climates potentially harmful to biodiversity and society are to be prevented.

Climate is a primary driver of biological processes, operating from individuals to ecosystems, and affects several aspects of human life. Therefore, climates without modern precedents could cause large and potentially serious impacts on ecological and social systems[1–5]. For instance, species whose persistence is shaped by the climate can respond by shifting their geographical ranges[4–7], remaining in place and adapting[5,8], or becoming extinct[8–11]. Shifts in species distributions and abundances can increase the risk of extinction[12], alter community structure[3] and disrupt ecological interactions and the functioning of ecosystems. Changing climates could also affect the following: human welfare, through changes in the supply of food[13] and water[14,15]; human health[16], through wider spread of infectious vector-borne diseases[17,18], through heat stress[19] and through mental illness[20]; the economy, through changes in goods and services[21,22]; and national security as a result of population shifts, heightened competition for natural resources, violent conflict and geopolitical instability[23]. Although most ecological and social systems have the ability to adapt to a changing climate, the magnitude of disruption in both ecosystems and societies will be strongly determined by the time frames in which the climate will reach unprecedented states[1,2]. Although several studies have documented the areas on Earth where unprecedented climates are likely to occur in response to ongoing greenhouse gas emissions[24,25], our understanding of climate change still lacks a precise indication of the time at which the climate of a given location will shift wholly outside the range of historical precedents.

To provide an indication of the projected timing of climate departure under alternative greenhouse gas emissions scenarios, we have developed an index that determines the year when the values of a given climatic variable exceed the bounds of historical variability for a particular location (Fig. 1a). We emphasize that although our index commonly identifies future dates, this does not imply that climate change is not already occurring. In fact, our index projects when ongoing climate change signals the start of a radically different climate. For this analysis we used the projections of 39 Earth System Models developed for the Coupled Model Intercomparison Project phase 5 (CMIP5). The bounds of climate variability were quantified as the minimum and maximum values yielded by the Earth System Models with the CMIP5 'historical' experiment, which for all models included the period from 1860 to 2005. This experiment included observed changes in atmospheric composition (reflecting both anthropogenic and natural sources) and was designed to model the climate's recent past and allow the validation of model outputs against available climate observations[26]. The year at which a climate variable moves out of the historical bounds was estimated independently with data from the Representative Concentration Pathways 4.5 (RCP45) and 8.5 (RCP85), which included the period from 2006 to 2100. These pathways or scenarios represent contrasting mitigation efforts between a concerted rapid CO_2 mitigation and a 'business-as-usual' scenario (CO_2 concentrations could increase to 538 and 936 p.p.m. by 2100, according to RCP45 and RCP85, respectively[27,28]). A more aggressive mitigation scenario (RCP 2.6) was not analysed, because it was not consistently used among models, and the implicit mitigation effort is considered currently unfeasible[29].

We analysed five climate variables for the atmosphere and two for the oceans (Extended Data Table 1). However, we report results with mean annual near-surface air temperature as our indicator of the climate, unless otherwise specified. Simulated and actual measurements of temperature for the period 1986–2005 were remarkably similar (Extended Data Fig. 1). In comparison with any individual model's results, the multi-model average best fitted the actual data (Extended Data Fig. 1). We therefore describe our results on the basis of multi-model averages. We show standard deviations to report the spatial variability in our results. Multi-model uncertainty (that is, the variability between models in the predicted years) was measured as the standard error of the mean, which for near-surface air temperature had a global value of 4.2 years for RCP45 and 2.7 years for RCP85 (Extended Data Fig. 2a). Multi-model uncertainty for all variables is shown in Extended Data Fig. 2.

[1]Department of Geography, University of Hawai'i at Mānoa, Honolulu, Hawai'i 96822, USA. [2]Department of Biology, University of Hawai'i at Mānoa, Honolulu, Hawai'i 96822, USA. [3]Hawai'i Institute of Marine Biology, University of Hawai'i at Mānoa, Kāne'ohe, Hawai'i 96744, USA. [4]Department of Oceanography, University of Hawai'i at Mānoa, Honolulu, Hawai'i 96822, USA. [5]Trans-disciplinary Organization for Subtropical Island Studies (TRO-SIS), University of the Ryukyus, Senbaru, Nishihara, Okinawa 903-0213, Japan.

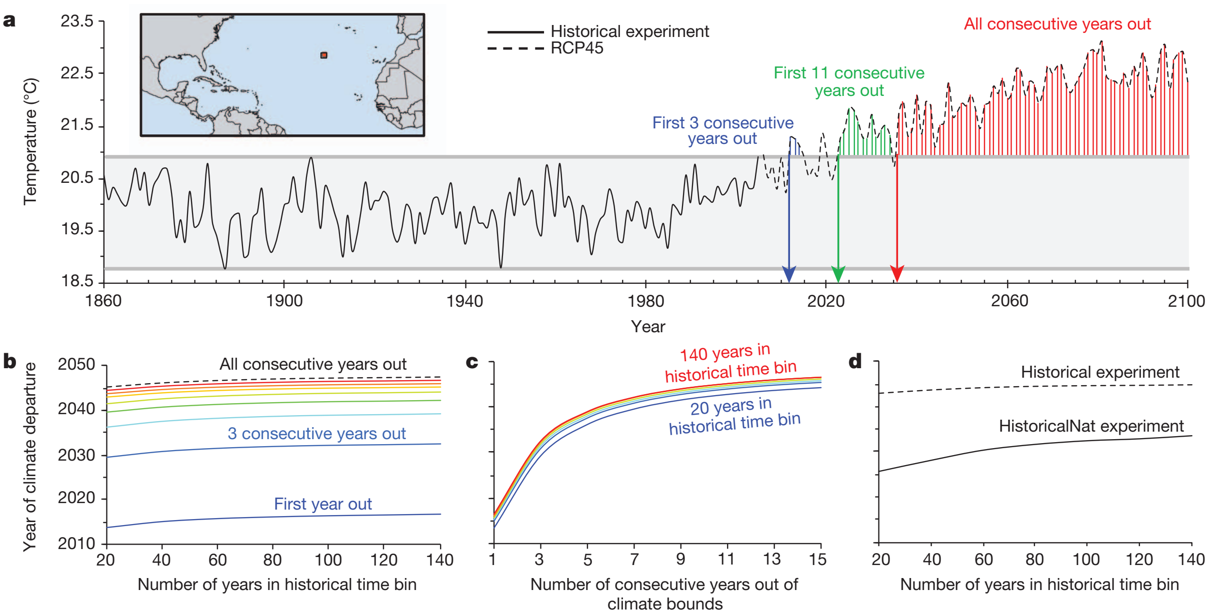


Figure 1 | Estimating the projected timing of climate departure from recent variability. **a**, Mean annual temperatures of an example grid cell (small square on map) exceed historical climate bounds (grey area) for three consecutive years starting in 2012 (blue arrow) and for 11 consecutive years after 2023 (green arrow); after 2036 (red arrow) all subsequent years remained outside the bounds (data from the Geophysical Fluid Dynamics Laboratory Earth System Model 2G). **b**, **c**, Effect of using different historical reference periods (**b**) and different numbers of consecutive years exceeding historical bounds (**c**) on the projected timing of climate departure from recent variability for global multi-model averages under RCP85. **d**, Comparison of the projected timing of climate departure from recent variability under RCP85, using the 'historical' and the 'historicalNat' experiments as reference to set the bounds of historical climate variability.

Sensitivity of the index

Three factors could affect the result of our index: first, the number of years used as the historical reference period (a longer historical reference period could yield broader bounds of climate variability); second, the number of consecutive years out of historical bounds in order to declare the timing of climate departure (for example, one single year out of historical bounds will probably occur sooner than several consecutive years); and third, the extent to which the historical reference period has been affected by anthropogenic greenhouse gases (the use of a period that already contains anthropogenic effects would yield broader bounds of climate variability).

To address the first two concerns we calculated the year when the climate exceeded the bounds of historical variability, using different historical reference periods and varying numbers of consecutive years out of the historical bounds. We found that the year in which the climate exceeded the bounds of historical variability changed minimally when using historical time bins ranging from 20 to 140 years (Fig. 1b). Despite the fact that our analysis was constrained by the 140 years of available model data, the observed relationship between historical time bins (X) and the year exceeding bounds of historical climate variability (Y) showed a strong correlation for both RCPs ($Y = 1.2773 \ln X + 2063.2$ for RCP45; $Y = 1.0865 \ln X + 2042.2$ for RCP85 (dashed line in Fig. 1b); $r^2 = 0.99$ and $P < 0.001$ in both cases). Under both RCPs, extrapolating these equations from a 140-year time bin to a 1,000-year time bin increased the estimated year exceeding the bounds of historical climate variability by only about 2 years. In contrast, the year at which the climate exceeded the bounds of historical variability was sensitive to the number of consecutive years out of historical bounds considered (Fig. 1c). As illustrated for the location in Fig. 1a, the climate will experience three consecutive years out of historical bounds by 2012, 11 consecutive years by 2023, and all consecutive years after 2036. To ensure robustness in our results we used the minimum and maximum values in the entire time series of the historical experiment from 39 Earth System Models available, suggesting that our results included the broadest historical climate bounds possible, given available data. Similarly, we used the first year after which all values would continuously exceed the bounds of historical climate variability.

To address the third concern we compared our results from the historical experiment with those obtained from an additional CMIP5 experiment, 'historicalNat', which simulated the same time span as the historical experiment but with only natural forcing (for example volcanoes and solar variability), while excluding anthropogenic greenhouse gas emissions[26]. The results of this analysis (Fig. 1d) indicated that the climate surpasses the bounds of historical variability about 18.5 years earlier under RCP45, and 11.5 years earlier under RCP85 when using historical simulations that excluded anthropogenic greenhouse-gas forcing (historicalNat experiment) compared with those that included it (historical experiment). We did not use the historicalNat experiment in our main results because it was available for only 17 out of 39 Earth System Models in CMIP5 (Extended Data Table 1) and this would have sacrificed the robustness obtained by using all available models. However, the sensitivity test above provides a quantitative measure of the likely adjustment in the projected timing of climate departure if historic periods without human effects were used to quantify past climate variability, and suggests that the reported values using the historical experiment are highly conservative.

The timing of climate departure

We found that the year at which the climate exceeds the bounds of historical variability depended on the modelled scenario. Under RCP45, the projected near-surface air temperature of the average location on Earth will move beyond historical variability by 2069 (±18 years s.d., 56 years in the future; solid blue line in Fig. 2d) and two decades earlier under RCP85 (that is, 2047 (±14 years s.d.), 34 years in the future; solid red line in Fig. 2d). These results are sobering indicators of the pace of climate change if one considers that the timing of climate departure will occur sooner if 'pristine' climate conditions (that is, the historicalNat experiment) are used to set historical climate bounds: 37.5 years in the future under RCP45, or 22.5 years under RCP85. For the ocean, the historical bounds of sea surface temperature will be surpassed on average by 2072 (±17 years s.d.) under RCP45 and by 2051 (±16 years s.d.) under RCP85 (Extended Data Fig. 3). When the index is calculated

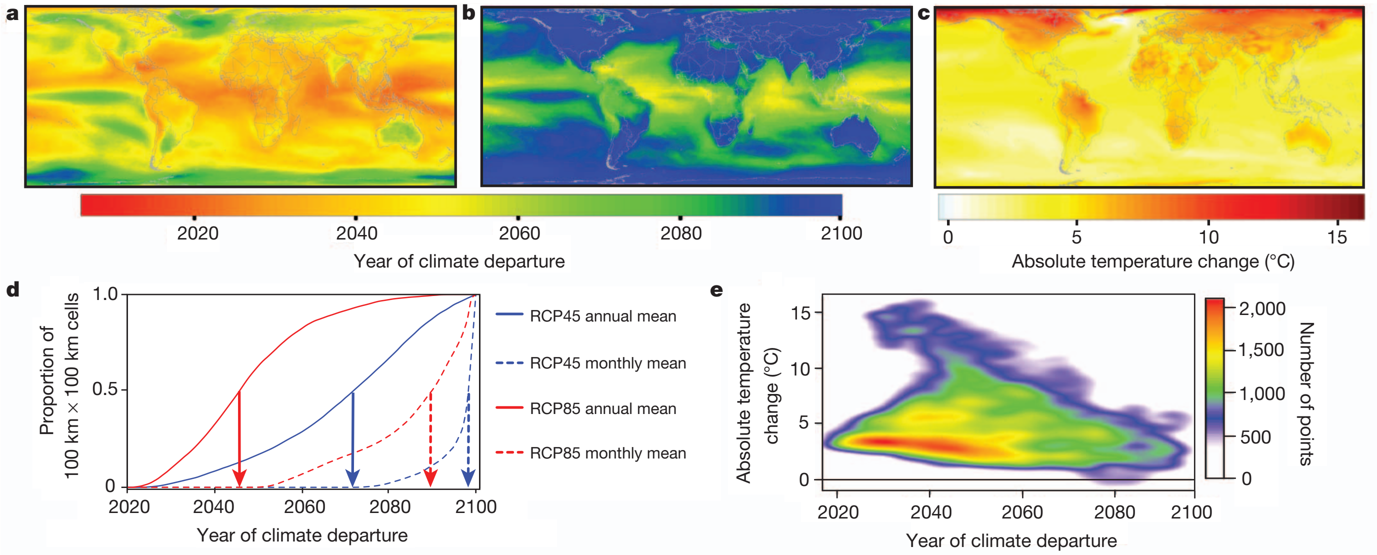


Figure 2 | The projected timing of climate departure from recent variability. **a**, **b**, Projected year when annual (**a**) or monthly (**b**) air temperature means move to a state continuously outside annual or monthly historical bounds, respectively. **c**, Absolute change in mean annual air temperature. (Results in **a**–**c** are based on RCP85.) **d**, Cumulative frequency of 100-km grid cells according to the projected timing of climate departure from recent variability for air temperature under two emissions scenarios (vertical lines indicate the median year). **e**, Scatter plot relating the grid cells from the map of absolute change (**c**) to the same grid cells from the map of projected timing of climate departure (**a**).

using monthly values (see Methods), all consecutive months will be out of the monthly historical bounds later in the century (Fig. 2b, dotted lines in Fig. 2d; see also Extended Data Fig. 3b, d). The estimated year when the climate exceeds historical variability is delayed when using monthly instead of annual averages, because one anomalous year is not necessarily caused by all months being extreme; thus, an anomalous year on average is likely to occur earlier than the year for which all months fall outside the monthly climate bounds. It is remarkable, however, that after 2050 most tropical regions will have every subsequent month outside of their historical range of variability (Fig. 2b). Although this is later than the yearly averages, we stress that this is an extreme situation, in which every month will be an extreme climatic record.

We used mean annual near-surface air temperature as the main proxy for the climate. However, other climate variables can change earlier or later than temperature in response to greenhouse gas emissions. To assess this effect, we analysed additional climate variables under both emissions scenarios and provided an overall climate assessment by estimating the year at which the first climate variable exceeded its historical bounds of variability. These variables included evaporation, transpiration, sensible heat flux and precipitation for the Earth's atmosphere, and surface pH for the ocean. For the atmosphere, the projected timing of climate departure did not change when considering other climate variables along with air temperature (Extended Data Fig. 4). This occurred because air temperature will experience the earliest and most sustained changes outside historical climate bounds in comparison with other atmospheric variables (that is, other atmospheric variables will continuously surpass their historical variability later than temperature; Extended Data Fig. 4). However, the projected timing of the ocean's climate departure was pushed forward to this decade when pH was considered alongside sea surface temperature. Global mean ocean pH moved outside its historical variability by 2008 ($\pm$3 years s.d.), regardless of the emissions scenario analysed (Extended Data Fig. 4). This result, which is consistent with recent studies[30], is explained by the fact that ocean pH has a narrow range of historical variability and that a considerable fraction of anthropogenic CO_2 emissions has been absorbed by the ocean[30,31].

Timing and absolute changes in climate

Absolute changes in the climate are often the means of detecting or assessing climate change and are expected to be considerably larger at higher latitudes (Fig. 2c; see also ref. 25). Measures of absolute changes in the climate have also dominated the dialogue on climate change (for example, avoiding 2 °C warming is a broadly recognized goal among scientists, policy makers and the public, because such change is forecast to generate deleterious consequences for society and the environment[25,32]). However, we found poor spatial correlation (Fig. 2e) between the absolute change in the climate expected by the year 2100 (Fig. 2c) and the year at which the climate would surpass historical precedents (Fig. 2a); this pattern was consistent among other climate variables (Extended Data Fig. 4). This result suggests that some aspects of climate change, which may be detrimental to biodiversity, are poorly accounted for by metrics of absolute changes in the climate; and implies that global biodiversity could face a climate change 'double jeopardy' from either large absolute changes or the early arrival of unprecedented climates.

We also found that the tropics will experience the earliest emergence of historically unprecedented climates (Fig. 2a, b). This probably occurs because the relatively small natural climate variability in this region of the world generates narrow climate bounds that can be easily surpassed by relatively small climate changes. However, small but fast changes in the climate could induce considerable biological responses in the tropics, because species there are probably adapted to narrow climate bounds[5,33–35]. This is a prime explanation for the decline in the range sizes of species towards lower latitudes (Rapoport's rule): having narrower tolerances, tropical species are largely restricted to the tropics; in comparison, the broader physiological tolerances of temperate species allow them to survive across a broader latitudinal span[33]. Furthermore, empirical and theoretical studies in corals[5,36,37], terrestrial ectotherms[34] and plants and insects[35] show that tropical species live in areas with climates near their physiological tolerances and are therefore vulnerable to relatively small climate changes.

The earliest emergence of unprecedented climates in the tropics and the limited tolerance of tropical species to climate change are troublesome results, because most of the world's biodiversity is concentrated in the tropics (Extended Data Fig. 5; see also ref. 33). We found that, on average, the projected timing of climate departure in marine and terrestrial biodiversity hotspots (*sensu* ref. 38, the top 10% most species-rich areas on Earth where a given taxon is found) will occur one decade earlier than the global average under either emissions scenario (Fig. 3). Coral hotspots will experience the earliest

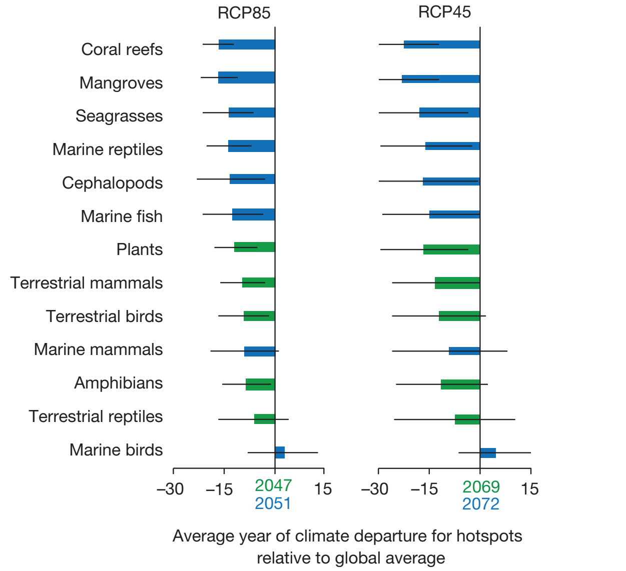


Figure 3 | The projected timing of climate departure from recent variability in global biodiversity hotspots. These plots indicate the difference between the average year in which the climate exceeds bounds of historical variability for each hotspot and the estimated global averages. The analysis was run independently for each hotspot, using mean annual air temperature for terrestrial taxa (green bars) or sea surface temperature for marine taxa (blue bars). Plots are centred at the respective global mean year for atmospheric (green numbers) and marine (blue numbers) environments. Horizontal bars rank the hotspots chronologically according to the mean year of unprecedented climates under RCP85. Horizontal black lines indicate the standard deviation among cells in the hotspots.

arrival of unprecedented climates: 2050 under RCP45 (about 23 years earlier than the global average), or 2034 under RCP85 (about 17 years earlier than the global average) (Fig. 3). With the exception of marine birds, whose hotspots are located at high latitudes (Extended Data Fig. 5d; see also ref. 39), unprecedented climates will occur at the latest by 2063 (RCP45) or 2042 (RCP85) in the hotspots of all other taxa considered (Fig. 3). Overall these results suggest that the overarching effect of climate change on biodiversity may occur not only as a result of the largest absolute changes in climate at high latitudes but also perhaps more seriously from small but prompt changes in the tropics. In short, the tropics will be highly vulnerable to climate change for at least three reasons: first, the earliest emergence of unprecedented climates will be there; second, tropical species are more vulnerable to small climate changes; and third, this region holds most of the Earth's species.

Discussion

The biological responses expected from the rapid emergence of historically unprecedented climates are likely to be idiosyncratic[40] and will depend on attributes such as species adaptive capacity, current genetic diversity, ability to migrate, current availability of habitats, disruption of ecological interactions, and ecological releases[40–44]. Although the extent of these responses in the future has been a topic of debate[45,46], considerable changes in community structure[3] and extinction[10] have been shown to have coincided with the emergence of unprecedented climates in the past. In addition, recent short-term extreme climatic events have been associated with die-offs in terrestrial[47–49] and marine[37] ecosystems, highlighting the potentially serious consequences of reaching historically unprecedented climates. Unfortunately, key conservation strategies such as protected areas, which may ameliorate the extent of several anthropogenic stressors, are unlikely to provide refuge from the expected effects of climate change, because protected areas within biodiversity hotspots will experience unprecedented climates at the same time as non-protected hotspot areas (Fig. 4a; see also ref. 50). The expansion and/or effectiveness of protected areas and other conservation

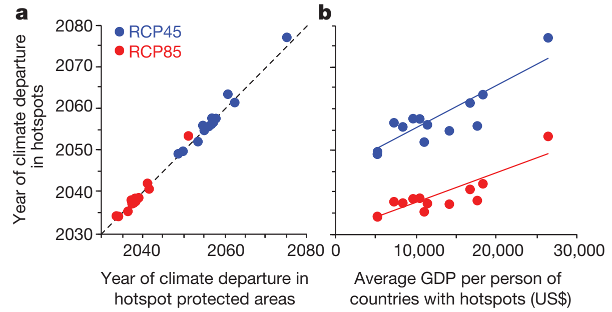


Figure 4 | Biodiversity hotspots: exposure to climate departures, and economic capacity to respond. Each point in these plots represents one of the 13 taxonomic biodiversity hotspots considered. **a**, Comparison of the year at which the climate exceeds bounds of historical variability between hotspots and protected areas in those hotspots. The dotted line shows the 1:1 relationship. **b**, Comparison of the year at which the climate exceeds bounds of historical variability in hotspots against the average Gross Domestic Product (GDP) per person for the countries encompassing the hotspots. The trend line for RCP45 is modelled with $y = 0.001x + 2045.2$ ($r^2 = 0.75$, $P < 0.05$; $n = 13$ hotspots) and for RCP85 with $y = 0.0007x + 2030.4$ ($r^2 = 0.75$, $P < 0.05$; $n = 13$ hotspots).

strategies could be further impaired by limited governmental capacity, because the earliest emergence of unprecedented climates will occur among hotspots predominantly located in low-income countries (Fig. 4b and Extended Data Fig. 6).

The emergence of unprecedented climates could also induce responses in human societies[1,2,13–16,19–22], and the resulting adjustments could be considerable because according to RCP45 roughly 1 billion people (about 5 billion people under RCP85) currently live in areas where climate will exceed historical bounds of variability by 2050 (Fig. 5a). The fact that the earliest climate departures occur in low-income countries (Fig. 5b) further highlights an obvious disparity between those who benefit economically from the processes leading to climate change and those who will have to pay for most of the environmental and social costs. This suggests that any progress to decrease the rate of ongoing climate change will require a bigger commitment from developed countries to decrease their emissions but will also require more extensive funding of social and conservation programmes in developing countries to minimize the impacts of climate change. Our results on the projected timing of climate departure from recent variability shed light on the urgency of mitigating greenhouse gas emissions if widespread changes in global biodiversity and human societies are to be prevented.

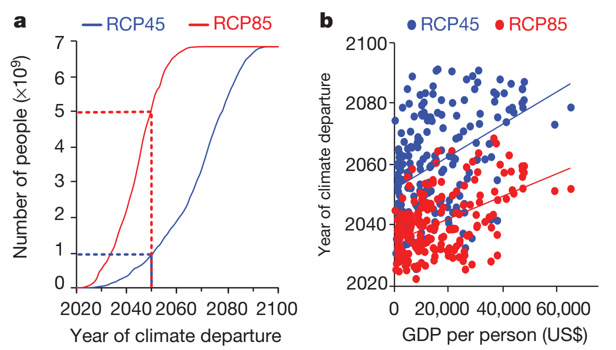


Figure 5 | Susceptibility of societies to climate departures, and economic capacity to respond. **a**, Plot of cumulative number of people against the years at which the climate of their current living areas will exceed historical climate bounds (dotted lines highlight the results for 2050). **b**, Relation between GDP per person and the average year of climate departure from recent variability. The trend line for RCP45 is modelled with $y = 0.0005x + 2051.9$ ($r^2 = 0.19$, $P < 0.05$; $n = 200$ countries or territories) and for RCP85 with $y = 0.0004x + 2034.5$ ($r^2 = 0.25$, $P < 0.05$; $n = 200$ countries or territories).

METHODS SUMMARY

We used the projections of seven climate variables (near-surface air temperature, sea surface temperature, precipitation, evaporation, transpiration, surface sensible heat flux, and ocean surface pH) from Earth System Models developed for CMIP5. As of March 2013 there were 39 Earth System Models from 21 climate centres in 12 countries that modelled at least one of the variables analysed (Extended Data Table 1). For each model and variable, we used the period 1860–2005 from the historical experiment (the longest time span common to all models), to establish the historical bounds of climate variability. The projections from RCP45 and RCP85, which simulate the period 2006–2100, were used to identify the year at which mean annual values of a given variable would exceed historical bounds. We also independently calculated the year at which the climate would exceed historical monthly variability by identifying the year containing the month after which all consecutive months, until 2100, exceeded monthly historical bounds (Fig. 2b). In total, for all variables and experiments, we processed 89,712 years of data comprising 1,076,544 monthly global maps, interpolated to an equal-area grid with a resolution of 100 km. Absolute climate change (Fig. 2c and Extended Data Figs 3c and 4) was calculated by subtracting contemporary averages (1996 to 2005) from future averages (2091 to 2100). Decadal averages were chosen to minimize aliasing by inter-annual variability. The projected timing of climate departure in biodiversity hotspots (Fig. 3), protected areas (Fig. 4), populated areas (Fig. 5a) and the economic capacity of countries (Fig. 5b) were estimated by sampling these areas from global map results on projected climate departure (based on near-surface air or sea surface temperature). All data sources are indicated in Extended Data Table 2.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 April; accepted 6 August 2013.

1. Reilly, J. & Schimmelpfennig, D. *Irreversibility, Uncertainty, and Learning: Portraits of Adaptation to Long-term Climate Change* (Springer, 2000).
2. Alley, R. B. *et al.* Abrupt climate change. *Science* **299,** 2005–2010 (2003).
3. Williams, J. W. & Jackson, S. T. Novel climates, no-analog communities and ecological surprises. *Front. Ecol. Environ* **5,** 475–482 (2007).
4. Peterson, A. T., Soberon, J., Pearson, R. G. & Martinez-Meyer, E. *Ecological Niches and Geographic Distributions* (Monographs in Population Biology Vol. 49) (Princeton Univ. Press, 2011).
5. Doney, S. C. *et al.* Climate change impacts on marine ecosystems. *Annu. Rev. Mar. Sci.* **4,** 11–37 (2012).
6. Parmesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421,** 37–42 (2003).
7. Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B. & Thomas, C. D. Rapid range shifts of species associated with high levels of climate warming. *Science* **333,** 1024–1026 (2011).
8. Parmesan, C. Ecological and evolutionary responses to recent climate change. *Annu. Rev. Ecol. Evol. Syst.* **37,** 637–669 (2006).
9. Thomas, C. D., Franco, A. M. A. & Hill, J. K. Range retractions and extinction in the face of climate warming. *Trends Ecol. Evol.* **21,** 415–416 (2006).
10. Crowley, T. J. & North, G. R. Abrupt climate change and extinction events in Earth history. *Science* **240,** 996–1002 (1988).
11. Mora, C. & Zapata, F. A. in *The Balance of Nature and Human Impact* (ed. K. Rohde) 239–257 (Cambridge Univ. Press, 2013).
12. Berg, M. P. *et al.* Adapt or disperse: understanding species persistence in a changing world. *Glob. Change Biol.* **16,** 587–598 (2010).
13. Lobell, D. B. & Gourdji, S. M. The influence of climate change on global crop productivity. *Plant Physiol.* **160,** 1686–1697 (2012).
14. Zhang, X. & Cai, X. Climate change impacts on global agricultural water deficit. *Geophys. Res. Lett.* **40,** 1111–1117 (2013).
15. Taylor, R. G. *et al.* Ground water and climate change. *Nature Clim. Change* **3,** 322–329 (2013).
16. Patz, J. A. & Olson, S. H. Climate change and health: global to local influences on disease risk. *Ann. Trop. Med. Parasitol.* **100,** 535–549 (2006).
17. Epstein, P. R. Climate change and infectious disease: stormy weather ahead? *Epidemiology* **13,** 373–375 (2002).
18. Khasnis, A. A. & Nettleman, M. D. Global warming and infectious disease. *Arch. Med. Res.* **36,** 689–696 (2005).
19. Sherwood, S. C. & Huber, M. An adaptability limit to climate change due to heat stress. *Proc. Natl Acad. Sci. USA* **107,** 9552–9555 (2010).
20. Berry, H., Bowen, K. & Kjellstrom, T. Climate change and mental health: a causal pathways framework. *Int. J. Public Health* **55,** 123–132 (2010).
21. Díaz, S., Fargione, J., Chapin, F. S. & Tilman, D. Biodiversity loss threatens human well-being. *PLoS Biol.* **4,** e277 (2006).
22. Tol, R. S. Estimates of the damage costs of climate change. Part 1. Benchmark estimates. *Environ. Resour. Econ.* **21,** 47–73 (2002).
23. Kloor, K. The war against warming. *Nature Rep. Clim. Change* **3,** 145–146, http://dx.doi.org/10.1038/climate.2009.120 (2009).
24. Williams, J. W., Jackson, S. T. & Kutzbach, J. E. Projected distributions of novel and disappearing climates by 2100 AD. *Proc. Natl Acad. Sci. USA* **104,** 5738–5742 (2007).
25. Solomon, S. *et al.* (eds). *Climate Change: The Physical Science Basis. Summary for Policymakers* (Cambridge Univ. Press, 2007).
26. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93,** 485–498 (2012).
27. Vuuren, D. P. *et al.* The representative concentration pathways: an overview. *Clim. Change* **109,** 5–31 (2011).
28. Meinshausen, M. *et al.* The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim. Change* **109,** 213–241 (2011).
29. van Vliet, J., den Elzen, M. G. & van Vuuren, D. P. Meeting radiative forcing targets under delayed participation. *Energy Econ.* **31,** S152–S162 (2009).
30. Raven, J. A. *et al.* (eds). *Ocean Acidification due to Increasing Atmospheric Carbon Dioxide* (Royal Society, 2005).
31. Zeebe, R. E., Zachos, J. C., Caldeira, K. & Tyrrell, T. Carbon emissions and acidification. *Science* **321,** 51–52 (2008).
32. Rockstrom, J. *et al.* A safe operating space for humanity. *Nature* **461,** 472–475 (2009).
33. Gaston, K. J. Global patterns in biodiversity. *Nature* **405,** 220–227 (2000).
34. Deutsch, C. A. *et al.* Impacts of climate warming on terrestrial ectotherms across latitude. *Proc. Natl Acad. Sci. USA* **105,** 6668–6672 (2008).
35. Colwell, R. K., Brehm, G., Cardelús, C. L., Gilman, A. C. & Longino, J. T. Global warming, elevational range shifts, and lowland biotic attrition in the wet tropics. *Science* **322,** 258–261 (2008).
36. Hoegh-Guldberg, O. Climate change, coral bleaching and the future of the world's coral reefs. *Mar. Freshw. Res.* **50,** 839–866 (1999).
37. Baker, A. C., Glynn, P. W. & Riegl, B. Climate change and coral reef bleaching: An ecological assessment of long-term impacts, recovery trends and future outlook. *Estuar. Coast. Shelf Sci.* **80,** 435–471 (2008).
38. Tittensor, D. P. *et al.* Global patterns and predictors of marine biodiversity across taxa. *Nature* **466,** 1098–1101 (2010).
39. Chown, S. L., Gaston, K. J. & Williams, P. H. Global patterns in species richness of pelagic seabirds: the Procellariiformes. *Ecography* **21,** 342–350 (1998).
40. La Sorte, F. A. & Jetz, W. Tracking of climatic niche boundaries under recent climate change. *J. Anim. Ecol.* **81,** 914–925 (2012).
41. Sorte, C. J. B. Predicting persistence in a changing climate: flow direction and limitations to redistribution. *Oikos* **122,** 161–170 (2013).
42. Devictor, V. *et al.* Differences in the climatic debts of birds and butterflies at a continental scale. *Nature Clim. Change* **2,** 121–124 (2012).
43. Zhu, K., Woodall, C. W. & Clark, J. S. Failure to migrate: lack of tree range expansion in response to climate change. *Glob. Change Biol.* **18,** 1042–1052 (2012).
44. Angert, A. L. *et al.* Do species' traits predict recent shifts at expanding range edges? *Ecol. Lett.* **14,** 677–689 (2011).
45. Baird, A. & Maynard, J. A. Coral adaptation in the face of climate change. *Science* **320,** 315–316 (2008).
46. Pandolfi, J. M., Connolly, S. R., Marshall, D. J. & Cohen, A. L. Projecting coral reef futures under global warming and ocean acidification. *Science* **333,** 418–422 (2011).
47. Allen, C. D. *et al.* A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *For. Ecol. Manage.* **259,** 660–684 (2010).
48. Carey, C. & Alexander, M. A. Climate change and amphibian declines: is there a link? *Divers. Distrib.* **9,** 111–121 (2003).
49. McKechnie, A. E. & Wolf, B. O. Climate change increases the likelihood of catastrophic avian mortality events during extreme heat waves. *Biol. Lett.* **6,** 253–256 (2010).
50. Mora, C. & Sale, P. Ongoing global biodiversity loss and the need to move beyond protected areas: a review of the technical and practical shortcomings of protected areas on land and sea. *Mar. Ecol. Prog. Ser.* **434,** 251–266 (2011).

Acknowledgements We thank D. Beilman for commenting on the paper; E. Wingert for help on the figures; D. Olsen for technical support; H. Kreft and the International Union for Conservation of Nature, BirdLife International, the Food and Agriculture Organization of the United Nations, the World Bank Database, the National Centers for Environmental Prediction, the World Database of Protected Areas, and the Gridded Human Population of the World Database for making their data openly available. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP5, and we thank the climate modelling groups (listed in Extended Data Table 1) for producing and making available their model outputs. This work was made possible by funding from the University of Hawai'i Sea Grant to C.M. The paper was developed as part of the graduate course on 'Methods for Large Scale Analyses' in the Department of Geography, University of Hawai'i at Mānoa. A.G.F. and T.W.G were supported by Pacific Islands Climate Change Cooperative (PICCC) award F10AC00077 and National Science Foundation Hawai'i EPSCoR grant no. EPS-0903833. R.J.L. was supported by the Pacific Islands Climate Science Center and PICCC award F10A00079. R.S.D. and E.J.T. were supported by National Science Foundation Graduate Fellowships, and I.F.-S. by a postdoctoral fellowship from the Japanese Society for the Promotion of Science.

Author Contributions All authors contributed equally to conceive the study, compile the data, conduct analyses, and write the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.M. (cmora@hawaii.edu).

METHODS

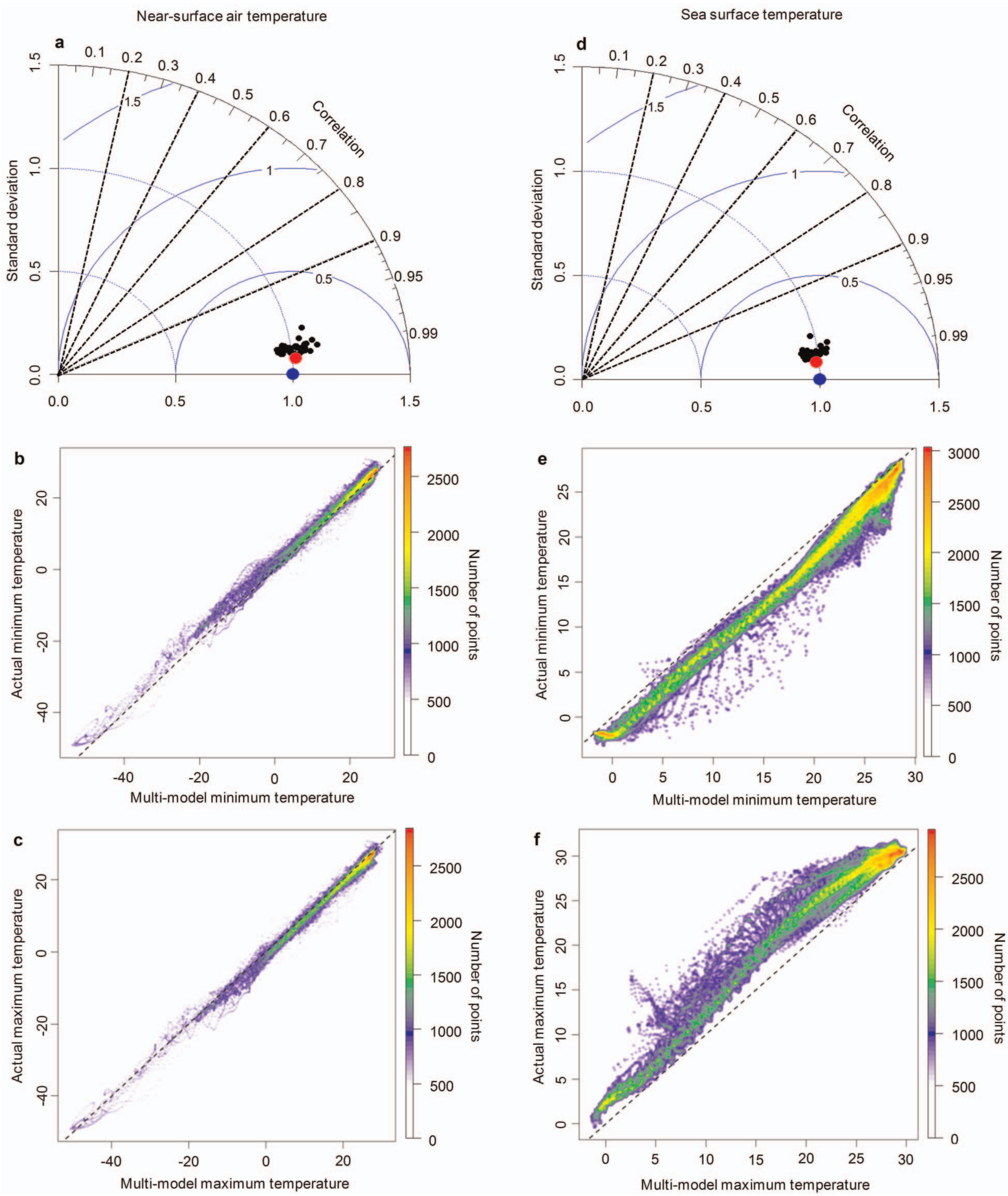
Approach. We used the projections of seven climatic variables (near-surface air temperature, sea surface temperature, precipitation, evaporation, transpiration, surface sensible heat flux, and ocean surface pH) from Earth System Models developed for CMIP5 (link to data source is provided in Extended Data Table 2). As of March 2013, there were 39 Earth System Models from 21 climate centres in 12 countries that modelled at least one of the variables analysed (Extended Data Table 1). We used the period 1860–2005 from the historical experiment (the longest time span common to all models) to establish the historical bounds of climate variability. The projections from RCP45 and RCP85, which simulate the period 2006–2100, were used to identify the year at which mean annual values of a given variable would exceed historical bounds (Fig. 1a). We also independently calculated the year at which the climate would exceed historical monthly variability by identifying the year containing the month after which all consecutive months, until 2100, exceeded monthly historical bounds. We used models that provided the complete data sets for the historical, RCP45 and RCP85 experiments. In total, for all variables and projections, we processed 89,712 years of data comprising 1,076,544 monthly global maps. Given that the CMIP5 models use different spatial grids to provide their data, we interpolated data from all models to an equal-area grid with a resolution of 100 km, using an area-weighted average interpolation method in which the data from the models' grid were transferred to our grid proportionally to the area that they occupied on our grid cells.

Sensitivity tests. We assessed three factors that could affect the result of our index: first, the number of years used as the historical reference period; second, the number of consecutive years out of historical bounds in order to declare the departure of climate; and third, the extent to which the historical reference period has been affected by anthropogenic greenhouse gases. To address these concerns we calculated the year when the climate exceeded the bounds of historical variability by using historical time bins varying from 20 to 140 years, by varying the number of consecutive years out of the historical bounds, and by comparing our results from the historical experiment with those obtained from an additional CMIP5 experiment, 'historicalNat', which simulated the same time span as the historical experiment, but with only natural forcing. On the basis of the results from these tests we decided to use, first, the minimum and maximum values of the entire historical period; second, the year after which all values were outside climate bounds, as our index; and third, the historical experiment as our reference period of climate variability, because it was available for all models in CMIP5 as opposed to the historicalNat experiment, which was available for only 17 out of 39 models. The test of our third concern, however, provides a measure of the difference in the results of our index using the historical reference period with human effects (historical experiment) compared with a historical reference period without human effects.

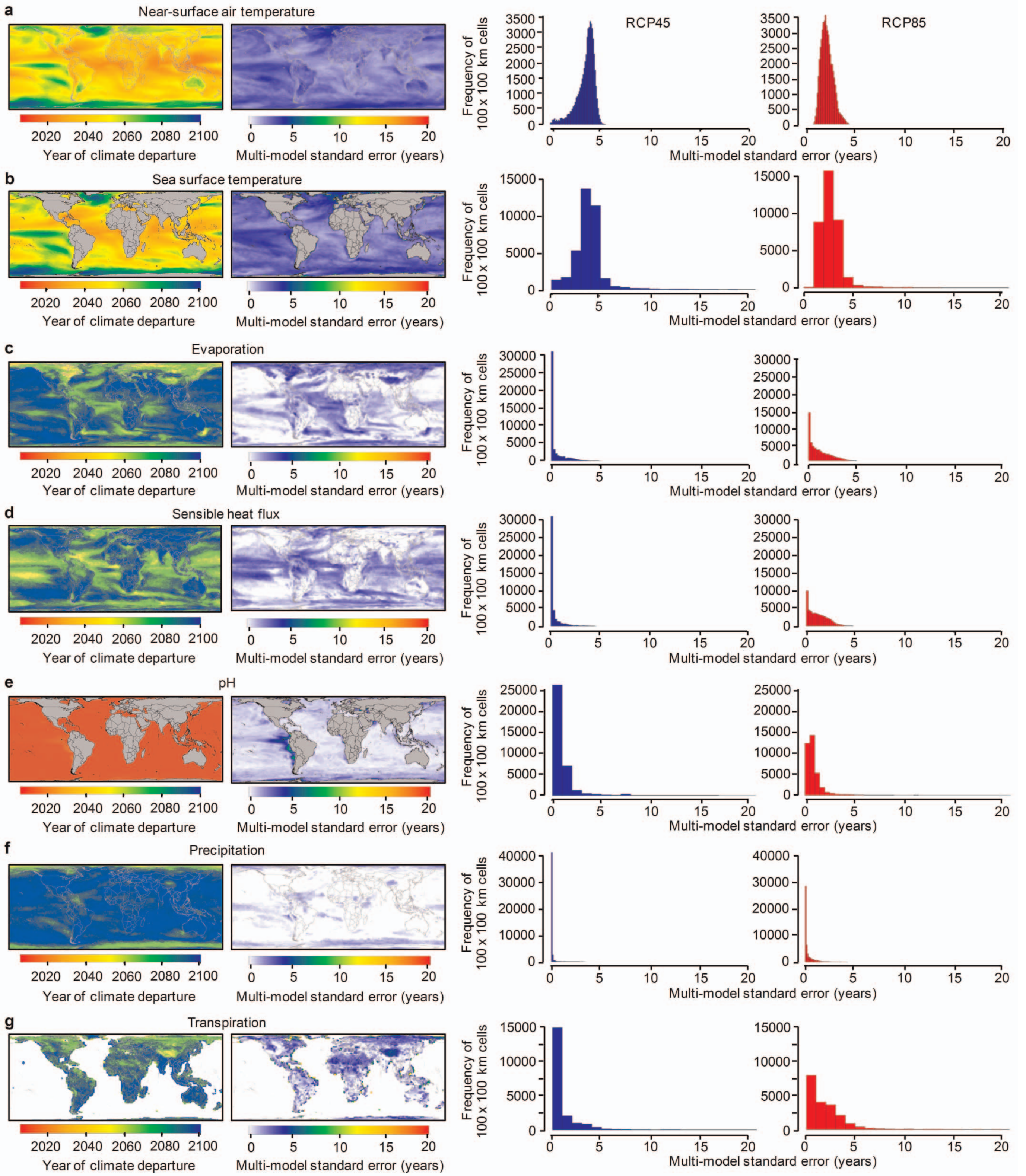
Determining the exposure of biodiversity and human societies. Biodiversity hotspots were outlined for 13 different marine (birds, cephalopods, corals, mammals, mangroves, marine fishes, reptiles and sea grasses) and terrestrial (amphibians, birds, mammals, reptiles and plants) taxonomic groups from global patterns of species richness (data sources are provided in Extended Data Table 2). Biodiversity hotspots were defined as the top 10% most species-rich areas on Earth where a given taxon is found (*sensu* ref. 33); their spatial distributions are outlined in Extended Data Fig. 5. We determined the exposure of global biodiversity hotspots to the projected timing of climate departure by intersecting biodiversity maps with maps of our index, and extracted the values in the overlapping cells. To assess the conservation potential of protected areas in biodiversity hotspots, we obtained data on protected areas, extracted all those in biodiversity hotspots and overlapped them with maps of our index to determine their timing of climate departure. We followed the same procedure for maps of population estimates and for GDP per person by country to assess the exposure of human societies and nations' economic capacities to the timing of climate departure (data sources are provided in Extended Data Table 2).

CMIP5 model robustness. We tested the robustness of CMIP5 Earth System Models by relating the simulated values and the multi-model average of near-surface air temperature and sea surface temperature from each model's output to 20 years of temperature observations (data sources are provided in Extended Data Table 2). We compared only air and sea surface temperature because of the broad availability and reliability of temperature observations. This comparison was restricted to the period 1986–2005, given the availability of actual data. The resulting metrics of fit include the correlation, the ratio of the standard deviations and the root mean squared error (all shown in Taylor diagrams in Extended Data Fig. 1a, d). Given our interest in historical climate bounds set by minimum and maximum climate values, we also compared the multi-model minimum and maximum projected values of temperature against actual minimum and maximum temperatures for the period 1986–2005 (Extended Data Fig. 1b, c, e, f).

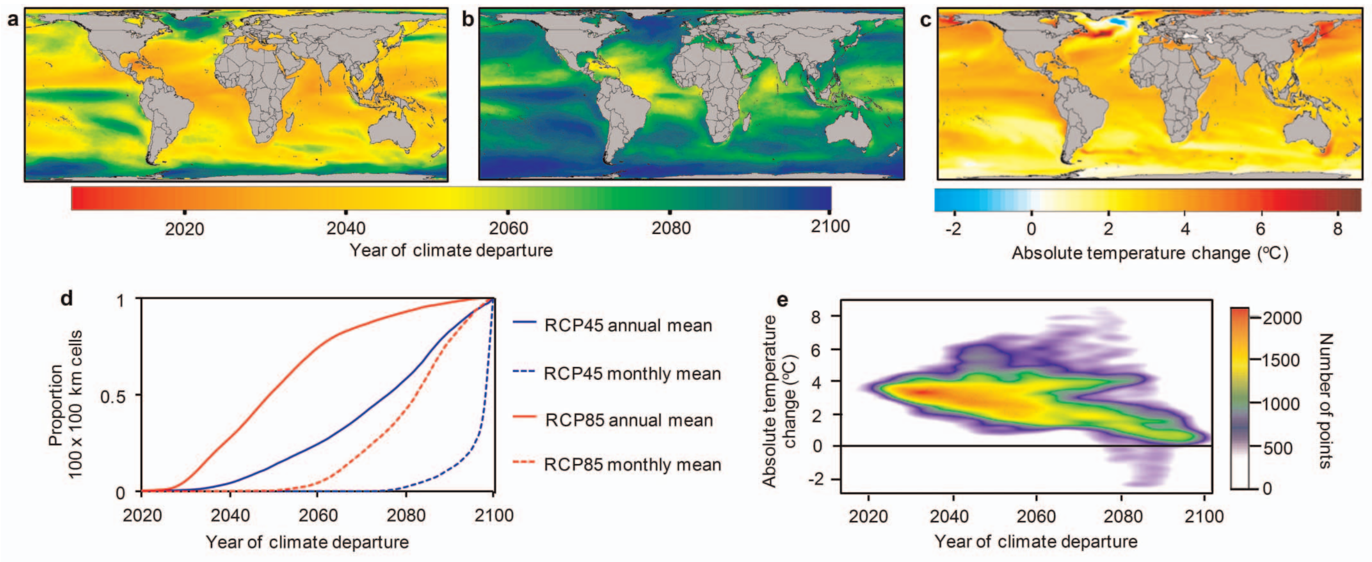
51. Kier, G. *et al.* A global assessment of endemism and species richness across island and mainland regions. *Proc. Natl Acad. Sci. USA* **106**, 9322–9327 (2009).



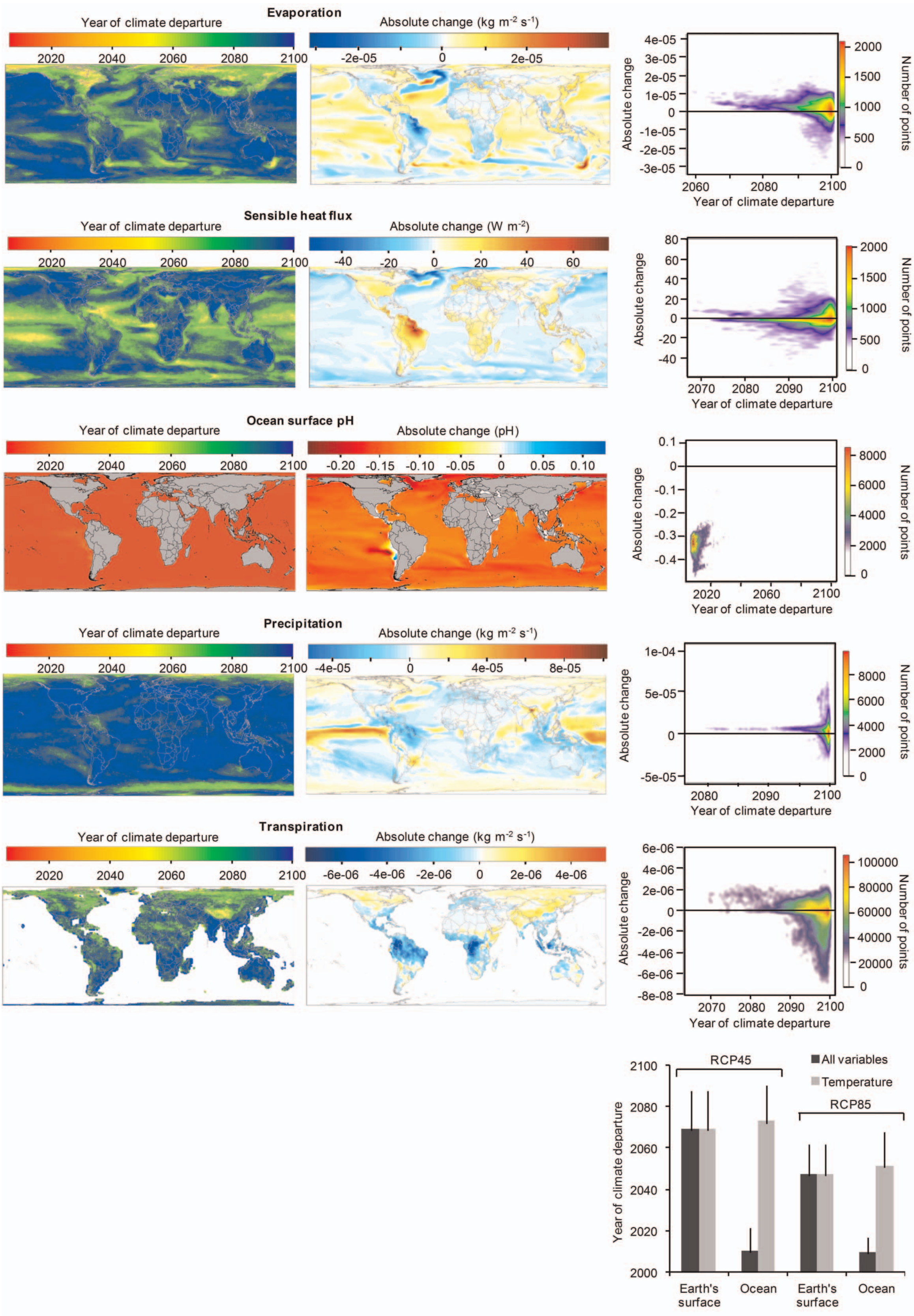
Extended Data Figure 1 | Evaluating robustness of Earth System Models. **a**–**c**, Analysis of near-surface air temperature. **d**–**f**, Analysis of sea surface temperature. **a**, **d**, Normalized Taylor diagrams. The Taylor diagrams compare actual observations with CMIP5 model simulations, and summarize three different metrics of similitude: the correlation (curved axis), the ratio of the standard deviations (*x* and *y* axes) and the root mean squared error (blue arcs). Blue points indicate perfect fit, red points the multi-model average, and black points the comparison of each Earth System Model to actual observations. The closer a red or black point is to the blue point, the better the fit between actual and simulated data. **b**, **c**, **e**, **f**, Comparison between actual and multi-model minimum (**b**, **e**) and maximum (**c**, **f**) temperatures for the 20-year period 1986–2005 (the time frame for which actual observations were mostly available). Dashed lines indicate the 1:1 relationship.



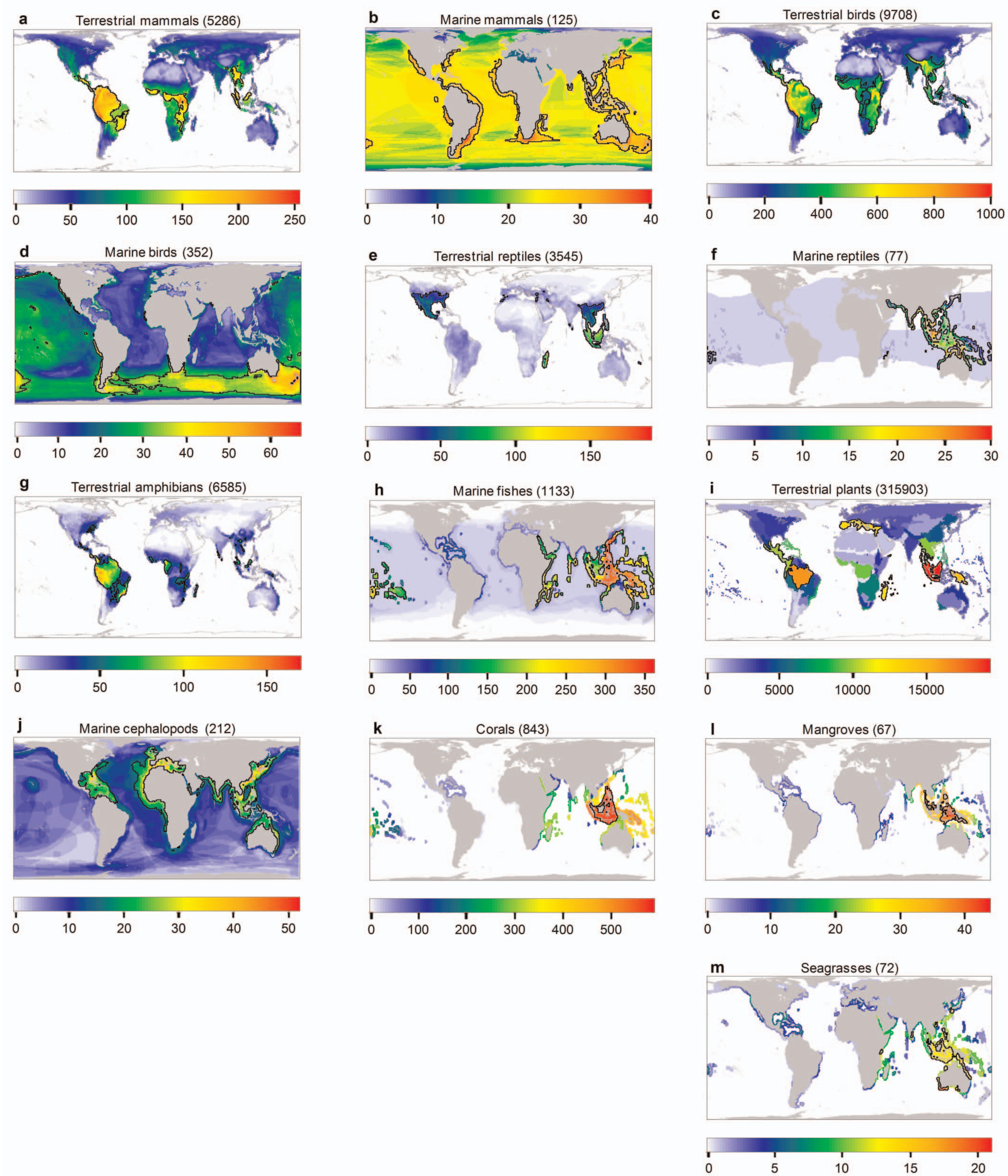
Extended Data Figure 2 | Multi-model uncertainty in the projected timing of climate departure. Results are shown for near-surface air temperature (**a**), sea surface temperature (**b**), evaporation (**c**), sensible heat flux (**d**), ocean surface pH (**e**), precipitation (**f**) and transpiration (**g**). Maps on the left show the mean year of climate departure under RCP85, and maps on the right illustrate the spatial patterns of inter-model standard error of the mean for RCP85. The histograms on the right indicate the frequency of grid cells by multi-model standard error of the mean according to each emissions scenario (blue, RCP45; red, RCP85).



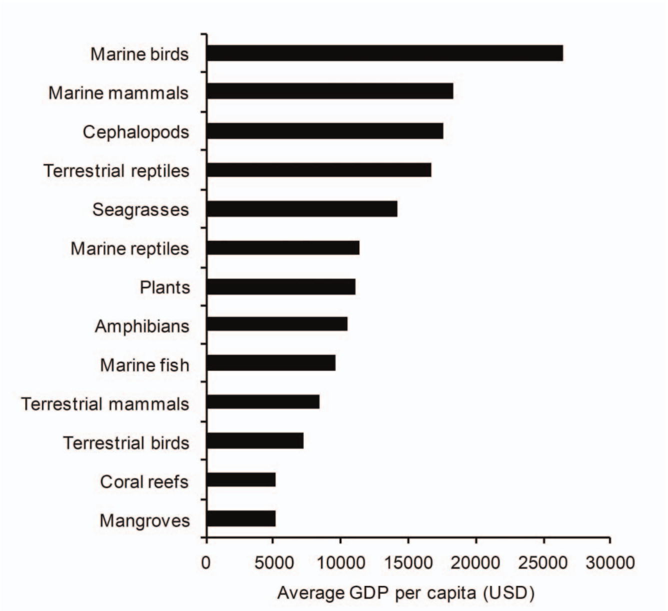
Extended Data Figure 3 | Sea surface temperature and projected timing of climate departure. **a**, **b**, Projected year when annual (**a**) or monthly (**b**) sea surface temperature means move to a state continuously outside annual or monthly historical bounds, respectively. **c**, Absolute change in mean annual sea surface temperature. (Results in **a**–**c** are based on RCP85.) **d**, Cumulative frequency of 100-km resolution grid cells according to the year of climate departure under the two emissions scenarios and for mean annual and monthly sea surface temperature. **e**, Scatter plot relating the grid cells from the absolute change map (**c**) to the same grid cells from the projected timing of climate departure map (**a**).



Extended Data Figure 4 | Projected timing of climate departure for different climate variables. We calculated the year of climate departure for five variables in addition to temperature. We considered the year of climate departure as the year at which the first variable exceeded its historical bounds of variability. The plots show the year of climate departure (left), the absolute change (middle) and the relation between the departure year and absolute change (right) under RCP85. The plot at the bottom right compares the global average year using temperature alone with the year when considering additional climate variables. Vertical lines indicate s.d.



Extended Data Figure 5 | Biodiversity hotspots. Global patterns of species richness were mapped for 13 marine and terrestrial taxa. For each taxon, we outlined biodiversity hotspots as the top 10% most species-rich places on Earth where the given taxon occurred (bold black lines). For mammals, birds, reptiles, amphibians, marine fishes, cephalopods, corals, mangroves and seagrasses (**a**–**h**, **j**–**m**), we used expert-verified geographical ranges to map patterns of species richness by counting the number of species whose ranges overlapped with an equal-area grid with a resolution of 100 km. **i**, For terrestrial vascular plants we used the number of species in different regions (data from ref. 51) and calculated species richness as the highest number of species occurring in the regions intersecting each 100-km resolution grid cell. The number of species or species richness used for each taxonomic group is indicated in parentheses.



Extended Data Figure 6 | Average gross domestic product (US$) per person for countries where the world's biodiversity hotspots are located. Horizontal bars represent the average GDP per person for the countries containing the hotspots for the 13 taxa examined.

Extended Data Table 1 | Earth System Models analysed

CENTER	COUNTRY	MODEL	tas	evspsbl	hfls	pr	tran	thetao	ph
Beijing Climate Center, China Meteorological Administration	China	BCC-CSM1.1*	✓	✓	✓	✓	✓	✓	
		BCC-CSM1.1(m)	✓	✓	✓	✓	✓	✓	
Canadian Centre for Climate Modelling and Analysis	Canada	CanESM2*	✓	✓	✓	✓	✓	✓	✓
Centro Euro-Mediterraneo per I Cambiamenti Clamatici	Italy	CMCC-CM	✓	✓	✓	✓		✓	
		CMCC-CMS	✓		✓	✓		✓	
Centre National de Recherches Meteorlogiques/Centre Européen de Recherche et Formation Avancée en Calcul Scientifique	France	CNRM-CM5*	✓	✓	✓	✓	✓	✓	
Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology	Australia	Access1.0	✓	✓	✓	✓		✓	
		Access1.3	✓	✓	✓	✓		✓	
Commonwealth Scientific and Industrial Research Organization with Queensland Climate Change Centre of Excellence	Australia	CSIRO-M3k3.6.0*	✓	✓	✓	✓		✓	
EC-EARTH Consortium	Europe	EC-EARTH	✓	✓	✓	✓		✓	
First Institute of Oceanography	China	FIO-ESM	✓	✓	✓	✓	✓	✓	
College of Global Change and Earth System Science, Beijing Normal Univ.	China	BNU-ESM*	✓	✓	✓	✓	✓	✓	
Institute for Numerical Mathematics	Russia	INM-CM4	✓	✓	✓	✓	✓	✓	
Institut Pierre-Simon Laplace	France	IPSL-CM5A-LR*	✓	✓	✓	✓	✓	✓	✓
		IPSL-CM5A-MR*	✓	✓	✓	✓	✓	✓	✓
		IPSL-CM5B-LR	✓	✓	✓	✓	✓	✓	✓
Laboratory for Atmospheric Sciences and Geophysical Fluid Dynamics and CESS Tsinghua University	China	FGOALS-g2	✓	✓	✓	✓		✓	
Laboratory for Atmospheric Sciences and Geophysical Fluid Dynamic, Chinese Academy of Sciences	China	FGOALS-s2			✓			✓	
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	Japan	MIROC5	✓	✓	✓	✓	✓	✓	
Japan Agency for Marine-Earth Science and Technology, Atmosphere and	Japan	MIROC-ESM*	✓	✓	✓	✓	✓	✓	
Ocean Research Institute and National Institute for Environmental Studies		MIROC-ESM-CHEM*	✓	✓	✓	✓	✓	✓	
Met Office Hadley Centre	UK	HadCM3				✓			
		HadGEM2-CC	✓	✓	✓	✓		✓	✓
		HadGEM2-ES*	✓	✓	✓	✓		✓	✓
Max Planck Institut für Meteorologie	Germany	MPI-ESM-LR	✓	✓	✓	✓	✓	✓	✓
		MPI-ESM-MR	✓	✓	✓	✓	✓	✓	✓
Meteorological Research Institute	Japan	MRI-CGCM3*	✓	✓	✓	✓	✓	✓	
NASA Goddard Institute for Space Studies	United States	GISS-E2-H*	✓	✓	✓	✓	✓	✓	
		GISS-E2-R*	✓	✓	✓	✓	✓	✓	
National Center for Atmospheric Research	United States	CCSM4*	✓	✓	✓	✓	✓	✓	
Norwegian Climate Centre	Norway	NorESM1-M*	✓	✓	✓	✓	✓	✓	
		NorESM1-ME	✓	✓	✓	✓	✓	✓	✓
National Institute of Meteorological Research/Korea Meteorological Administration	Korea	HadGEM2-AO	✓		✓	✓		✓	
NOAA Geophysical Fluid Dynamics Laboratory	United States	GFDL-CM3*	✓	✓	✓	✓	✓	✓	
		GFDL-ESM2G	✓	✓	✓	✓	✓	✓	✓
		GFDL-ESM2M*	✓	✓	✓	✓	✓	✓	✓
National Science Foundation	United States	CESM1(BGC)	✓	✓	✓	✓	✓	✓	✓
		CESM1(CAM5)	✓	✓	✓	✓	✓	✓	
		CESM1(WACCM)	✓	✓	✓	✓	✓	✓	

The table shows the list of models used for each variable analysed. We considered only models that provided the complete series of data from 1860 to 2100 under the historical, RCP45 and RCP85 experiments; asterisks indicate that the model also provided results for the 'historicalNat' experiment (data source shown in Extended Data Table 2). The variables analysed included near-surface air temperature (CMIP5 variable name 'tas', in K), precipitation ('pr', $kg\,m^{-2}\,s^{-1}$), evaporation ('evspsbl', $kg\,m^{-2}\,s^{-1}$), transpiration ('tran', $kg\,m^{-2}\,s^{-1}$), surface upward sensible heat flux ('hfss', $W\,m^{-2}$), surface sea water potential temperature ('thetao', K) and pH ('ph', $mol\,H\,kg^{-1}$).

Extended Data Table 2 | Data sources

AGENCY	SOURCE
Earth System Models Data	
World Climate Research Programme*	http://cmip-pcmdi.llnl.gov/cmip5/availability.html
Biodiversity Data	
Expert verified ranges	
BirdLife International and NatureServe 2011	www.birdlife.org/
International Union for Conservation of Nature 2012	http://www.iucnredlist.org/technical-documents/spatial-data
Food and Agriculture Organization of the United Nations 2005	www.fao.org
Species richness	
Kier et al.	Kier G, Kreft H, Lee TM, Jetz W, Ibisch PL, Nowicki C, Mutke J, Barthlott W (2009) A global assessment of endemism and species richness across island and mainland regions. Proceedings of the National Academy of Sciences., 106(23): 9322-9327.
Protected Areas	
World Database on Protected Areas	http://protectedplanet.net/
Gridded Observational Data	
NOAA Earth System Research Laboratory	http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html http://www.esrl.noaa.gov/psd/cgi-bin/db_search/DBSearch.pl?Dataset=NCEP+GODAS
Socio-Economic Data	
Gridded Human Population of the World Database	http://sedac.ciesin.columbia.edu/data/collection/gpw-v3
World Bank Database	http://data.worldbank.org/

* We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP5, and we thank the climate modeling groups (listed in Table S1 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

doi:10.1038/nature12617

A Silurian placoderm with osteichthyan-like marginal jaw bones

Min Zhu[1], Xiaobo Yu[1,2], Per Erik Ahlberg[3], Brian Choo[1], Jing Lu[1], Tuo Qiao[1], Qingming Qu[3], Wenjin Zhao[1], Liantao Jia[1], Henning Blom[3] & You'an Zhu[1]

The gnathostome (jawed vertebrate) crown group comprises two extant clades with contrasting character complements. Notably, Chondrichthyes (cartilaginous fish) lack the large dermal bones that characterize Osteichthyes (bony fish and tetrapods). The polarities of these differences, and the morphology of the last common ancestor of crown gnathostomes, are the subject of continuing debate. Here we describe a three-dimensionally preserved 419-million-year-old placoderm fish from the Silurian of China that represents the first stem gnathostome with dermal marginal jaw bones (premaxilla, maxilla and dentary), features previously restricted to Osteichthyes. A phylogenetic analysis places the new form near the top of the gnathostome stem group but does not fully resolve its relationships to other placoderms. The analysis also assigns all acanthodians to the chondrichthyan stem group. These results suggest that the last common ancestor of Chondrichthyes and Osteichthyes had a macromeric dermal skeleton, and provide a new framework for studying crown gnathostome divergence.

The early fossil record of gnathostomes is dominated by four seemingly well-defined groups: the Chondrichthyes and Osteichthyes, which are still extant, and the extinct Placodermi and Acanthodii[1–9]. Osteichthyans and placoderms have macromeric dermal skeletons dominated by large bony plates[4,7], but these skeletons have usually been regarded as non-homologous a priori and given different nomenclatures[10], because of perceived fundamental differences between placoderm and osteichthyan dermal bone patterns[1,2,7,11]. Chondrichthyans and acanthodians have micromeric dermal skeletons composed principally of scales. The current consensus is that placoderms are members of the gnathostome stem group[11–16] and most probably form a paraphyletic stem segment[15,16]. Recent phylogenetic analyses[15,16] have also recovered acanthodians as a paraphyletic assemblage spanning the chondrichthyan and osteichthyan stem groups and the crownward end of the gnathostome stem group (Fig. 1). This phylogenetic pattern implies that the macromeric dermal skeleton of placoderms was replaced by a micromeric condition in the common ancestor of crown gnathostomes, with subsequent *de novo* acquisition of a non-homologous macromeric skeleton in osteichthyans. However, the presence of placoderm-like features in the earliest osteichthyans (for example, dermal pelvic girdles[17] and multipartite dermal shoulder girdles with spinal plates[18–20]) suggests a conservation of pattern between the placoderm and osteichthyan macromeric dermal skeletons[4,21–23]. Homology of these skeletons would imply that the crown gnathostome node condition includes a macromeric dermal skeleton, and that the micromeric condition of acanthodians and chondrichthyans is derived (Fig. 1). The resolution of this issue hinges on the phylogenetic placement of an array of problematic fossil taxa, many yielding limited anatomical data.

Here we present a new fish from the Silurian of China[20,24,25] that combines a placoderm-like dermal skull roof, braincase and shoulder girdle with osteichthyan-like dermal bones of the mandibular and hyoid arches[2] and a palatoquadrate sharing derived characteristics with crown gnathostomes[12]. As this character combination has never been observed before, our description below uses placoderm terms for the skull roof and trunk armour[26], but osteichthyan terms for the dermal bones of the mandibular and hyoid arches (Supplementary Table 1).

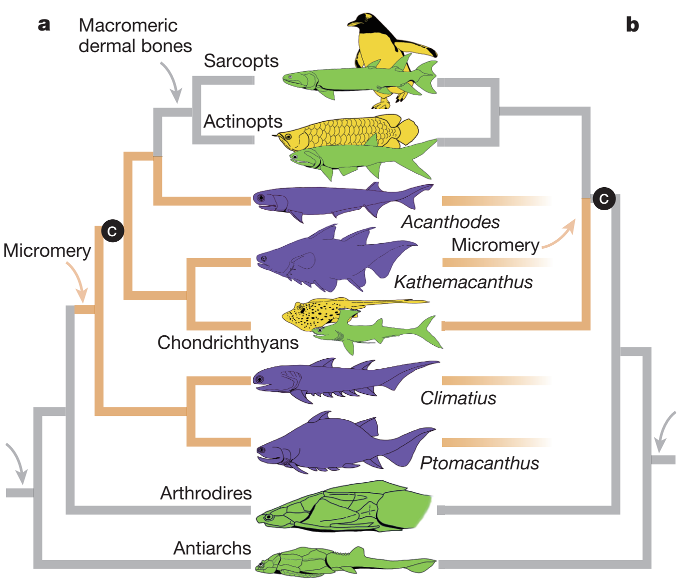


Figure 1 | Competing hypotheses of dermal skeleton condition at the crown gnathostome node. a, Simplified gnathostome phylogeny based on ref. 16. The micromeric condition (brown branches) in acanthodians (purple) and chondrichthyans brackets the crown gnathostome node (C). The macromeric condition (grey branches) is proposed as non-homologous in osteichthyans (sarcopterygians and actinopterygians) and placoderms (for example, arthrodires and antiarchs). **b,** Inferred macromeric condition at the crown gnathostome node (C), as suggested by recent findings supporting placoderm-osteichthyan dermal skeleton homology. This implies that micromery in acanthodians and chondrichthyans is derived, and questions the positions of acanthodians as stem gnathostomes and stem osteichthyans.

[1]Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, China. [2]Department of Biological Sciences, Kean University, Union, New Jersey 07083, USA. [3]Subdepartment of Evolution and Development, Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, Uppsala 752 36, Sweden.

Systematic palaeontology

Gnathostomata Gegenbaur, 1874

Entelognathus primordialis gen. et sp. nov.

Etymology. The generic name derives from the Greek *enteles* (complete) and *gnathos* (jaw), referring to the complete set of dermal marginal jaw bones. The specific name is from the Latin *primordialis* (primordial).

Holotype. IVPP V18620, an articulated fish with head shield and trunk armour (Fig. 2a–d and Supplementary Figs 13–17 and 21a, b).

Referred material. Skulls, V18621.1-15; sclerotic rings, V18622.1-2; cheek–palatoquadrate complexes, V18622.3-10; trunk armour plates, V18622.11-30.

Locality and horizon. Xiaoxiang Reservoir, Qujing, Yunnan, China; Kuanti Formation. Late Ludlow, Silurian, *Ozarkodina snajdri* Conodont Zone, 419 million years ago[20,24,25].

Diagnosis. A jawed stem gnathostome combining placoderm-like skull roof and trunk armour with osteichthyan-like marginal jaw bones (premaxilla, maxilla and dentary) and operculogular series. Skull roof lacking preorbital plates, premedian plate ventrally positioned between rostral and premaxilla, main lateral line meeting postmarginal line in anterior paranuchal plate, small orbital fenestra enclosed by a large oblong tripartite sclerotic ring, and anterior lateral plate with deep posterior notch for pectoral fenestra.

Description. The holotype represents a three-dimensionally preserved fish with articulated head shield and trunk armour, approximately 11 cm long, suggesting a total body length of over 20 cm.

Skull roof and sclerotic ring

The skull roof is flat mesially and strongly arched laterally. Pineal, central and nuchal plates cover most of the braincase along the midline, flanked by paired postorbital, marginal, and anterior and posterior paranuchal plates along the trajectory of the infraorbital and main lateral line grooves. The large hexagonal central plate seems to have a single ossification centre, whereas most placoderms have paired centrals[1,2,7]. Its location anterior to the nuchal aligns *Entelognathus* with arthrodires, in which paired centrals meeting in the midline occupy the corresponding position. In acanthothoracids and petalichthyids, the nuchal extends forwards medially to separate the centrals[3,4,26]. The posterior margin of the nuchal is overlapped by a partially exposed short bone, here termed 'postnuchal' (Supplementary Figs 11 and 18a–c), which resembles the 'extrascapular' of the arthrodire *Sigaspis*[27] and the petalichthyid *Eurycaraspis*[28]. The main lateral line joins the postmarginal line in the anterior paranuchal, not in the marginal as in other placoderms. The groove for the supraorbital line, with its associated postnasal and preorbital plates, is absent. No foramen for the endolymphatic duct or the pineal organ is visible.

The short and narrow snout comprises a rostral, a premedian and premaxillae (Figs 2, 3, 4i–k and 5, Supplementary Fig. 20 and Supplementary Video) that appear to be fused mesially. The rostral bends ventrally, giving the snout a near-vertical profile. A single pair of subtriangular nasal openings lies between the rostral and the sclerotic ring. The premaxilla has a narrow facial lamina, and a broad palatal lamina that flares out posteriorly and articulates with the palatal lamina of the maxilla. The facial lamina anteriorly forms a conjoined flush surface with the premedian and posteriorly contacts the lacrimal and the maxilla. The tubercles on the palatal lamina are smaller than those on the facial lamina, and somewhat tooth-like in shape. No separate vomer (or anterior supragnathal of placoderms) is present.

The small orbital fenestra is enclosed by a large oblong sclerotic ring, which comprises three intimately fused sclerotic plates (Figs 2, 3 and 4e, i–k) as in antiarchs[29]. The ring has sutural contacts to the adjacent skull roof bones, with an anterior process embraced by the premedian and premaxilla.

Dermal cheek, operculogular and jaw bones

The cheek complex comprises the jugal (placoderm suborbital), lacrimal, maxilla and a possible quadratojugal (Fig. 3), and has a sensory canal pattern broadly resembling that of arthrodires[23,30].

The elongate jugal has loose anterior contacts with the sclerotic ring and lacrimal, but its ventral suture with the maxilla is so tight that the bones are difficult to distinguish except by their ornament patterns. The infraorbital canal enters the jugal dorsally, runs ventrally to join the supraoral canal, and then continues anteroventrally via the slender lacrimal to terminate below the orbit. The supraoral canal runs posteroventrally to enter the facial lamina of the maxilla.

The maxilla has a slender facial lamina and a broad palatal lamina (Fig. 4a–d and Supplementary Figs 21, 22). At a level corresponding to the anterior margin of the fossa for adductor mandibular muscle, the ventral margin of the maxilla presents a sharp vertical bend. Anterior

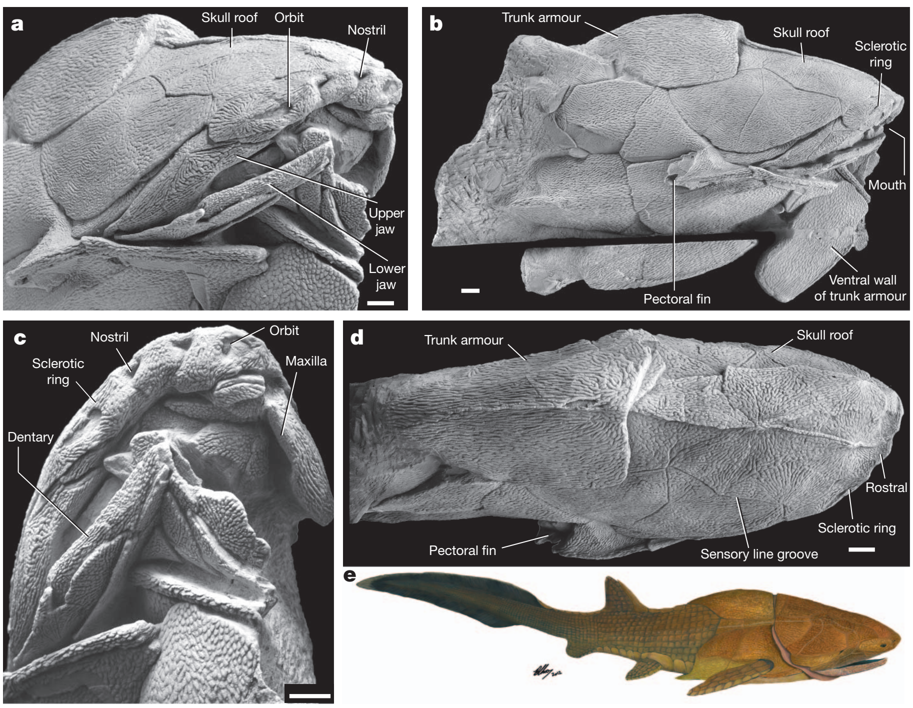


Figure 2 | *Entelognathus primordialis* gen. et sp. nov., a 419-million-year-old jawed fish from the Kuanti Formation (Late Ludlow, Silurian), Qujing, Yunnan. **a**–**d**, Holotype V18620, a three-dimensionally preserved specimen with head and trunk armour in anterolateral (**a**), lateral (**b**), anteroventral (**c**) and dorsal (**d**) views. A small part of the left trunk armour was accidentally sawed off as extraneous material and repositioned in **b**. Scale bars, 1 cm. **e**, Life restoration.

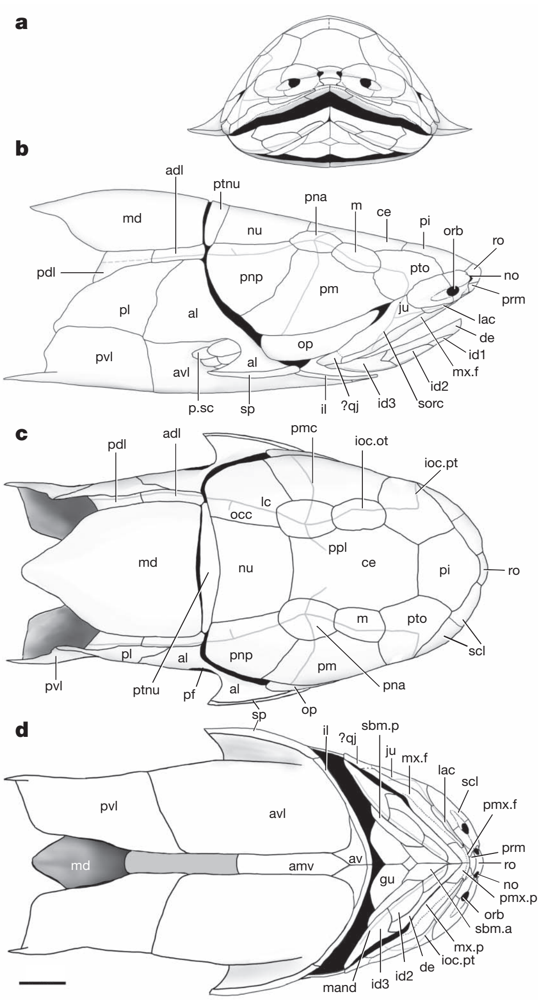


Figure 3 | ***Entelognathus primordialis* gen. et sp. nov. a–d**, Restoration of the dermal skeleton in anterior (**a**), lateral (**b**), dorsal (**c**) and ventral (**d**) views. Scale bars, 1 cm. adl, anterior dorsolateral plate; al, anterior lateral plate; amv, anterior medioventral plate; av, anteroventral plate; avl, anterior ventrolateral plate; ce, central plate; de, dentary; gu, principal gular; id1–3, first to third infradentary; il, interolateral plate; ioc.ot, otic branch of infraorbital line groove; ioc.pt, postorbital branch of infraorbital line groove; ju, jugal; lac, lacrimal; lc, main lateral line groove; m, marginal plate; mand, mandibular line groove; md, median dorsal plate; mx.f; facial lamina of maxilla; mx.p, palatal lamina of maxilla, no, nostril; nu, nuchal plate; occ, occipital cross commissure; op, opercular; orb, orbital fenestra; pdl, posterior dorsolateral plate; pf, pectoral fenestra; pi, pineal plate; pl, posterior lateral plate; pm, postmarginal plate; pmc, postmarginal line groove; pmx.f, facial lamina of premaxilla; pmx.p, palatal lamina of premaxilla, pna, anterior paranuchal plate; pnp, posterior paranuchal plate; ppl; posterior pitline; prm, premedian; ptnu, postnuchal plate; pto, postorbital plate; pvl, posterior ventrolateral plate; p.sc; pectoral fin scales; qj, quadratojugal; ro, rostral plate; sbm.a, anterior submandibular; sbm.p, posterior submandibular; scl, sclerotic plate; sorc, supraoral line groove; sp, spinal plate.

to this bend, the palatal lamina extends forward to join the palatal lamina of the premaxilla, forming a continuous horizontal shelf inward of the jaw margin, against which the mandible occludes. Elongate tubercles on the facial lamina of the maxilla and nearby dermal bones grade into close-packed rounded tubercles on the horizontal shelf. There seems to be no teeth or cusps along the oral margin. No separate dermal bones (dermopalatines, ectopterygoids or posterior supragnathals of placoderms) are found medial to the maxilla.

A small, roughly triangular ossification, tentatively identified as a quadratojugal, lies at the posterior tip of the cheek complex, and is firmly sutured with the adjacent maxilla and jugal. The position of this bone resembles that of a placoderm postsuborbital. X-ray tomography shows that the partially exposed opercular (placoderm submarginal) is an elongate crescentic bone with an opercular cartilage attached to its inner surface (Figs 2 and 3 and Supplementary Fig. 18).

The mandible (Figs 2a–c, 3a, b, d and 4i–j and Supplementary Fig. 13) is covered externally by a relatively short, edentulous dentary and three infradentaries. The dentary has a deep facial lamina but only a narrow edge dorsally, with no mesial lamina corresponding to the palatal lamina of the maxilla. No separate dermal bones (coronoids and prearticular, or infragnathals of placoderms) are found on the mesial surface of the mandible. X-ray tomography shows the mandibular line groove running through infradentary 2, and confirms the presence of the submandibular and gular series in the holotype (Fig. 4i and Supplementary Fig. 23). Medial to the infradentaries, the elongate anterior and posterior submandibulars as well as a lozenge-shaped median gular are preserved *in situ*, whereas the paired principal gulars are anteriorly displaced. Overall, this arrangement recalls the pattern seen in the lungfish *Scaumenacia*[2].

Trunk armour

The long trunk armour resembles that of early arthrodires such as a phlyctaeniid[30]. The anterior dorsolateral plate bears a flat anterior flange (Supplementary Fig. 26c). Coupled with the lack of an articular fossa or condyle on the posterior margin of the paranuchal plate, this flange indicates a sliding dermal neck joint as in actinolepids, phyllolepids and *Wuttagoonaspis*[3,26]. The anterior lateral plate (Supplementary Fig. 15a) has extensive overlap surfaces for the skull roof and the opercular, and a deep posterior notch for the pectoral fenestra. The narrow spinal and the adjoining plates form a triangular cutwater immediately anterior to the pectoral fin. The posterior ventrolateral plates have pointed tips protruding posterolaterally. Their medial margins diverge posteriorly and lack overlap surfaces, indicating that the ventral side of the trunk armour may have been incompletely enclosed posteriorly.

Braincase

The perichondrally ossified braincase (Fig. 5) is broad and flattened, resembling that of early arthrodires[23,30]. Anteriorly, the ectethmoid process lies slightly posteroventral to the nasal capsule. The orbital cavity is posteriorly separated from the postorbital fossa by a pila-like structure (Fig. 5a, c, d). Behind it lies a developed postorbital fossa. The anterior postorbital process carries an oval unfinished articular facet for the hyomandibula. Medial to this facet is a small foramen for the hyomandibular trunk of the facial nerve. The otico-occipital portion makes up about two thirds of the braincase length. In contrast to ptyctodonts with three endocranial ossifications[31], *Entelognathus* resembles other placoderms in lacking basicranial and otico-occipital fissures (Fig. 5a and Supplementary Fig. 19a, c, e). Behind the posterior postorbital process, the braincase tapers to form a large cucullaris fossa, and carries a lateral groove for the jugular vein. There is a strong posterolaterally-oriented craniospinal process, but no supravagal process.

Palatoquadrate

The perichondrally ossified palatoquadrates of partly disarticulated skulls (Fig. 5a and Supplementary Fig. 19e) show that the ethmoid articulation abuts the ectethmoid process of the braincase. The jaw articulation comprises a prearticular process and a quadrate concavity (Fig. 4b–d), as in acanthodians and chondrichthyans[2], but unlike the bipartite convex articulation in osteichthyans[32]. The metapterygoid region of the palatoquadrate is expanded ventrally into a commissural lamina that confines the adductor musculature mesially (Figs 4b–d and 6b), as in crown gnathostomes[12]. X-ray tomography of an isolated cheek-palatoquadrate complex (Fig. 4f, g and Supplementary Fig. 22e–h) reveals the adductor chamber. Rather than being wide open anteriorly as in osteichthyans (where the metapterygoid and autopalatine attachments of the palatoquadrate on the cheek are completely

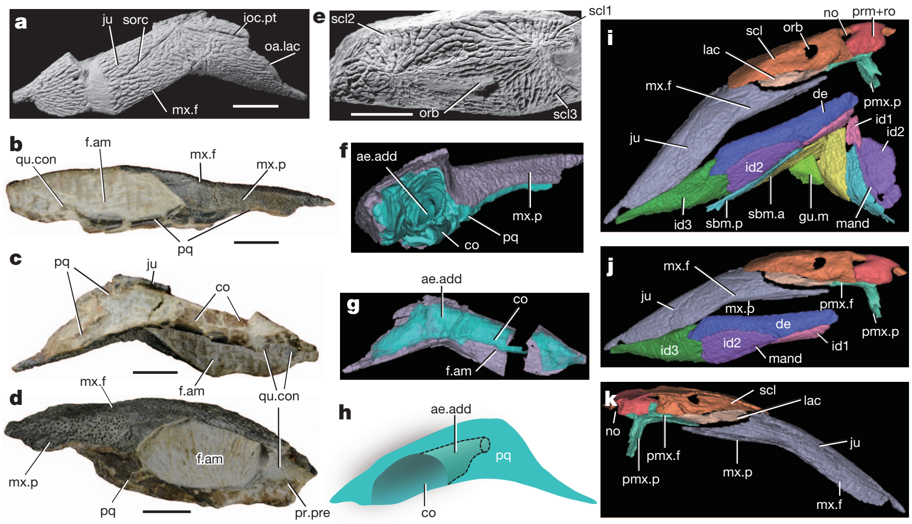


Figure 4 | *Entelognathus primordialis* gen. et sp. nov. a–c, Right upper jaw in external (**a**), ventral (**b**) and medioventral (**c**) views, V18622.3. **d**, Left upper jaw in ventral view, V18622.4. **e**, Disarticulated sclerotic ring, V18622.1. **f**, **g**, Computerized tomography restorations of an isolated cheek-palatoquadrate complex in posterior (**f**) and medioventral (**g**) views, V18622.3; the palatoquadrate rendered semi-transparent in **g** to show the anterior tunnel-like extension of the adductor chamber. **h**, Restoration of the palatoquadrate in lateral view. **i**–**k**, Computerized tomography restorations of the holotype V18620 showing the lower and upper jaws and neighbouring bones. Right side in anteroventral (**i**, **j**) and internal (**k**) views. Scale bars, 5 mm. ae.add, anterior extension of adductor chamber; co, commissural lamina; f.am, fossa for adductor mandibulae muscle; gu.m, median gular; oa.lac, overlapped area by lacrimal; pq, palatoquadrate; pr.pre, prearticular process; qu.con, quadrate concavity; scl1–3, first to third sclerotic plate. Other abbreviations as in Fig. 3.

separate), it has a tunnel-like anterior extension that traverses the palatoquadrate and emerges on its mesial face. A similar configuration exists in many arthrodires including *Buchanosteus*[33] and probably *Dicksonosteus*[30].

Phylogenetic relationship

To explore the phylogenetic position of *Entelognathus* and the impact of its characters on gnathostome phylogeny, we conducted analyses using a modified version of the data set of ref. 16 with 253 characters

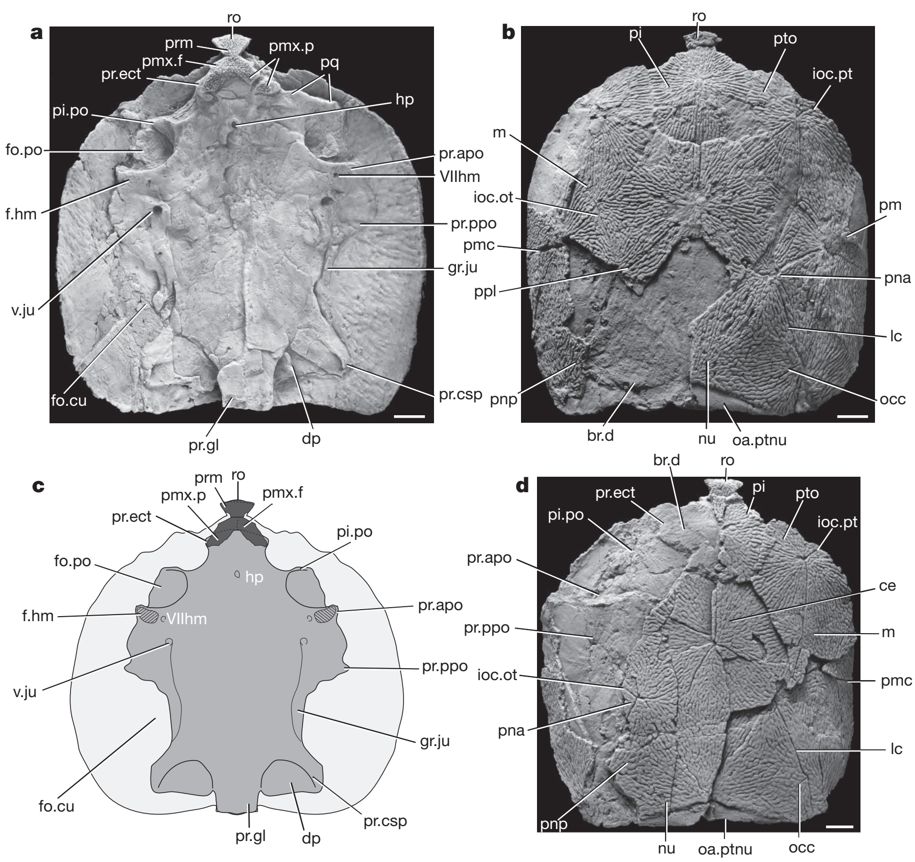


Figure 5 | Skull roof and braincase of *Entelognathus primordialis* gen. et sp. nov. a–b, Skull in dorsal (**a**) and ventral (**b**) views, V18621.1. **c**, Restoration of the skull in ventral view. **d**, Skull in dorsal view, with part of skull roof removed to show the dorsal wall of braincase, V18621.2. Scale bars, 5 mm. br.d, dorsal wall of braincase; dp, depression in occipital portion of braincase; fo.cu, cucullaris fossa; fo.po, postorbital fossa; f.hm, hyomandibular facet; gr.ju, groove for jugular vein; hp, hypophysial opening; oa.ptnu, overlapped area by postnuchal plate; pi.po, postoribital pila; pr.apo, anterior postorbital processs; pr.csp, craniospinal process; pr.ect, ectethmoid process; pr.gl, occipital glenoid process; pr.ppo, posterior postorbital process; v.ju, opening for jugular vein; VIIhm, hyomandibular trunk of facial nerve. Other abbreviations as in Figs 3 and 4.

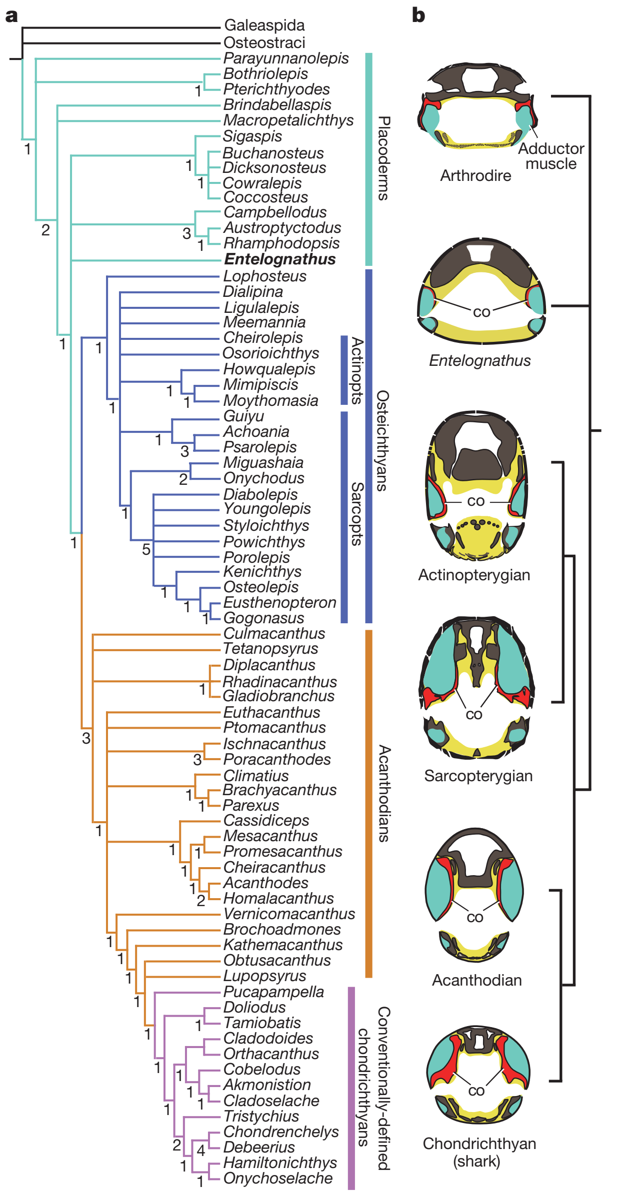


Figure 6 | Results of phylogenetic analysis and palatoquadrate conditions among major gnathostome groups. **a**, Strict consensus of 1,117 most parsimonious trees resulting from a modified data set. *Entelognathus* is placed in a polychotomy with arthrodires, ptyctodonts and crown gnathostomes, whereas all acanthodian taxa fall on the chondrichthyan stem. Numbers at nodes represent Bremer decay indices. **b**, Diagrammatic reconstruction of head in transverse section through the fossa for adductor mandibulae muscle to show the relationships of the braincase, palatoquadrate (red) and adductor muscle (green). *Entelognathus* resembles crown gnathostomes in having a developed commissural lamina (co). Schematic drawings modified from ref. 12.

and 75 taxa (see Supplementary Information and Methods). The strict consensus of 1,117 most parsimonious trees (MPTs) places *Entelognathus* in a polychotomy with arthrodires, ptyctodonts and crown gnathostomes (Fig. 6a), whereas the 50% majority-rule consensus favours *Entelognathus* as the sister group of crown gnathostomes (Supplementary Fig. 2a). Notably, our result differs from those of refs 15 and 16 in placing all acanthodians on the chondrichthyan stem, although it agrees with them with regard to placoderm paraphyly, osteichthyan monophyly and the internal topology of conventionally defined chondrichthyans.

Homology of macromeric skeletons

Because the analysis of ref. 16 placed some acanthodians in the osteichthyan and gnathostome stem groups, the phenetic similarities between acanthodians and chondrichthyans were interpreted as symplesiomorphies. Our phylogenetic result indicates that the similarities actually represent synapomorphies of the chondrichthyan total group; examples include body scales with a neck and bulging base, and a skull roof comprising undifferentiated plates or tesserae (Fig. 6a). Placement of all acanthodians in the chondrichthyan stem group also implies macromery in the last common ancestor of crown gnathostomes. Although partial to complete loss of macromery in acanthodians and chondrichthyans is implicit both in our phylogeny and those of refs 15 and 16, as a corollary of placoderm paraphyly, the important difference is whether osteichthyan macromery is a novelty or a retained primitive feature. The partly micromeric condition in the cheek and jaws of the stem osteichthyan *Dialipina*[34] could be interpreted as evidence for *de novo* evolution of macromery, but its skull roof pattern resembles those of both arthrodires and crown osteichthyans (Supplementary Fig. 10), suggesting homology across the gnathostome crown group node. Micromeric regions also occur between the large skull plates of some placoderms (for example, *Gemuendina*[1,2]), and in the snouts of many sarcopterygians (for example, *Powichthys*[2]).The significance of these occurrences needs to be investigated further, as does the broader question of placoderm-osteichthyan skeletal pattern homology (Supplementary Information).

Dermal jaw and palatoquadrate evolution

Until now, dermal marginal jaw bones (premaxilla, maxilla and dentary) were regarded as key synapomorphies of Osteichthyes[32,35], and have never been found in any placoderms. The similarity in shape and topographic correspondence between the marginal jaw bones of *Entelognathus* and osteichthyans strongly suggest their homology. However, unlike the question of macromery versus micromery, dermal jaw homology cannot be resolved unambiguously from our phylogenetic analysis. Three positions are recovered for *Entelognathus*, as sister group to crown gnathostomes (66% of MPTs), in a clade with arthrodires and ptyctodonts (28% of MPTs), and as sister group to ptyctodonts plus crown gnathostomes (6% of MPTs) (Supplementary Fig. 4). Only in the first topology does one of two equally parsimonious character optimizations identify the dermal jaw bones of *Entelognathus* as homologous with those of osteichthyans. This lack of resolution stems partly from gaps in the data set and partly from incongruent character distributions among key taxa, notably between *Entelognathus*, ptyctodonts and crown gnathostomes.

If the most frequently resolved position of *Entelognathus* as the sister group of crown gnathostomes (Supplementary Fig. 2a) is accepted as a working hypothesis, it offers a new perspective on the early evolution of crown gnathostome morphology. *Entelognathus* combines osteichthyan-like marginal jaw bones with a conservative placoderm-like condition of the braincase, dermal skull roof and trunk armour, whereas the palatoquadrate shows similarities with that of crown gnathostomes. In other placoderms, the metapterygoid portion is low, and there is at most a very narrow commissural lamina[33], whereas crown gnathostomes have a tall metapterygoid (making the palatoquadrate 'cleaver-shaped') and a developed commissural lamina confining the adductor muscle mesially[12] (Fig. 6b). In *Entelognathus*, a developed commissural lamina is present, but the metapterygoid portion is low. The low metapterygoid is linked to the possession of a broad arthrodire-like braincase with low side walls[3,23], contrasting with the narrower and deeper braincase of crown gnathostomes. Such a character complement suggests that changes to the muscular organization and biting surfaces preceded the emergence of a cleaver-shaped palatoquadrate at or just before the crown gnathostome node.

METHODS SUMMARY

To determine the phylogenetic position of *Entelognathus*, and to evaluate the existing gnathostome phylogenies, we conducted phylogenetic analyses (parsimony approach and Bayesian inference) using a modified data set with 253 characters and 75 taxa. Parsimony-based results place *Entelognathus* in a polychotomy with

arthrodires, ptyctodonts and crown gnathostomes. Unexposed or partially exposed structures of the holotype and a cheek-palatoquadrate complex were examined using X-ray micro-computerized tomography. The specimens in Figs 2, 4a, e and 5 were coated with ammonium chloride sublimate.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 January; accepted 29 August 2013.

Published online 25 September 2013.

1. Moy-Thomas, J. A. & Miles, R. S. *Palaeozoic Fishes* (Chapman & Hall, 1971).
2. Jarvik, E. *Basic Structure and Evolution of Vertebrates* Vol. 1 (Academic, 1980).
3. Goujet, D. F. Placoderm interrelationships: a new interpretation, with a short review of placoderm classifications. *Proc. Linn. Soc. N. S. W.* **107,** 211–243 (1984).
4. Gardiner, B. G. The relationship of placoderms. *J. Vertebr. Paleontol.* **4,** 379–395 (1984).
5. Maisey, J. G. Heads and tails: a chordate phylogeny. *Cladistics* **2,** 201–256 (1986).
6. Schultze, H.-P. in *The Skull* Vol. 2 (eds Janke, J. & Hall, B. K.) 189–254 (Univ. Chicago Press, 1993).
7. Janvier, P. *Early Vertebrates* (Oxford Univ. Press, 1996).
8. Young, G. C. Placoderms (armored fish): dominant vertebrates of the Devonian period. *Annu. Rev. Earth Planet. Sci.* **38,** 523–550 (2010).
9. Long, J. A. *The Rise of Fishes: 500 Million Years of Evolution* 2nd edn (John Hopkins Univ. Press, 2011).
10. Gross, W. Peut-on homologuer les os des Arthrodires et des Téléostomes? *Colloq. Int. CNRS* **104,** 69–74 (1962).
11. Young, G. C. The relationships of placoderm fishes. *Zool. J. Linn. Soc.* **88,** 1–57 (1986).
12. Schaeffer, B. in *Problèmes actuels de Paléontologie-Evolution des Vertébrés* Vol. 218 (ed. Lehman, J. P.) 101–109 (Actes du Colloque International CNRS, 1975).
13. Goujet, D. F. in *Major Events in Early Vertebrate Evolution: Palaeontology, Phylogeny, Genetics and Development* (ed. Ahlberg, P. E.) 209–222 (Taylor & Francis, 2001).
14. Johanson, Z. Vascularization of the osteostracan and antiarch (Placodermi) pectoral fin: similarities, and implications for placoderm relationships. *Lethaia* **35,** 169–186 (2002).
15. Brazeau, M. D. The braincase and jaws of a Devonian 'acanthodian' and modern gnathostome origins. *Nature* **457,** 305–308 (2009).
16. Davis, S. P., Finarelli, J. A. & Coates, M. I. *Acanthodes* and shark-like conditions in the last common ancestor of modern gnathostomes. *Nature* **486,** 247–250 (2012).
17. Zhu, M. *et al.* Fossil fishes from China provide first evidence of dermal pelvic girdles in osteichthyans. *PLoS ONE* **7,** e35103 (2012).
18. Zhu, M., Yu, X.-B. & Janvier, P. A primitive fossil fish sheds light on the origin of bony fishes. *Nature* **397,** 607–610 (1999).
19. Zhu, M. & Schultze, H.-P. in *Major Events in Early Vertebrate Evolution: Palaeontology, Phylogeny, Genetics and Development* (ed. Ahlberg, P.) 289–314 (Taylor & Francis, 2001).
20. Zhu, M. *et al.* The oldest articulated osteichthyan reveals mosaic gnathostome characters. *Nature* **458,** 469–474 (2009).
21. Gardiner, B. G. The relationships of the palaeoniscid fishes, a review based on new specimens of *Mimia* and *Moythomasia* from the Upper Devonian of western Australia. *Bull. Br. Mus. Nat. Hist. Geol.* **7,** 173–428 (1984).
22. Forey, P. L. & Gardiner, B. G. Observations on *Ctenurella* (Ptyctodontida) and the classification of placoderm fishes. *Zool. J. Linn. Soc.* **86,** 43–74 (1986).
23. Stensiö, E. A. Anatomical studies on the arthrodiran head. Part 1. Preface, geological and geographical distribution, the organization of the head in the Dolichothoraci, Coccosteomorphi and Pachyosteomorphi. Taxonomic appendix. *Kungl. Svenska Vetenskap. Hand.* **9,** 1–419 (1963).
24. Zhao, W.-J. & Zhu, M. Siluro-Devonian vertebrate biostratigraphy and biogeography of China. *Palaeoworld* **19,** 4–26 (2010).
25. Zhu, M., Liu, Y.-H., Jia, L.-T. & Gai, Z.-K. A new genus of eugaleaspidiforms (Agnatha:Galeaspida) from the Ludlow, Silurian of Qujing, Yunnan, southwestern China. *Vertebr. PalAsiat.* **50,** 1–7 (2012).
26. Denison, R. H. in *Handbook of Paleoichthyology* Vol. 2 (ed. Schultze, H.-P.) 1–128 (Gustav Fischer, 1978).
27. Goujet, D. F. *Sigaspis*, un nouvel arthrodire du Dévonien inférieur du Spitsberg. *Palaeontogr. A* **143,** 73–88 (1973).
28. Liu, Y.-H. in *Early Vertebrates and Related Problems of Evolutionary Biology* (eds Chang, M.-M., Liu, Y.-H. & Zhang, G.-R.) 139–177 (Science Press, 1991).
29. Burrow, C. J., Newman, M. J., Davidson, R. G. & den Blaauwen, J. L. Sclerotic plates or circumorbital bones in early jawed fishes? *Palaeontology* **54,** 207–214 (2011).
30. Goujet, D. F. *Les Poissons Placodermes du Spitsberg. Arthrodires Dolichothoraci de la Formation de Wood Bay (Dévonien Inférieur)* (CNRS, 1984).
31. Trinajstic, K., Long, J. A., Johanson, Z., Young, G. & Senden, T. New morphological information on the ptyctodontid fishes (Placodermi, Ptyctodontida) from western Australia. *J. Vertebr. Paleontol.* **32,** 757–780 (2012).
32. Friedman, M. & Brazeau, M. D. A reappraisal of the origin and basal radiation of the Osteichthyes. *J. Vertebr. Paleontol.* **30,** 36–56 (2010).
33. Young, G. C. New information on the structure and relationships of *Buchanosteus* (Placodermi: Euarthrodira) from the Early Devonian of New South Wales. *Zool. J. Linn. Soc.* **66,** 309–352 (1979).
34. Schultze, H.-P. & Cumbaa, S. L. in *Major Events in Early Vertebrate Evolution: Palaeontology, Phylogeny, Genetics and Development* (ed. Ahlberg, P. E.) 315–332 (Taylor & Francis, 2001).
35. Botella, H., Blom, H., Dorka, M., Ahlberg, P. E. & Janvier, P. Jaws and teeth of the earliest bony fishes. *Nature* **448,** 583–586 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M.-M. Chang, J. Long, G. Young, D. Goujet and J.-Q. Wang for discussions, X.-F. Lu, J. Zhang and C.-H. Xiong for specimen preparation, Y.-M. Hou for X-ray micro-computerized tomography. This work was supported by the Major State Basic Research Projects (2012CB821902) of MST of China, and the National Natural Science Foundation of China (40930208). P.E.A. and Q.Q. were supported by ERC Advanced Investigator Grant 233111 and a Wallenberg Scholarship from the Knut and Alice Wallenberg Foundation, both awarded to P.E.A.

Author Contributions M.Z. conceived the project. M.Z., W.Z., L.J., Y.Z., J.L. and T.Q. did the fieldwork. M.Z., P.E.A., T.Q., J.L., Q.Q., B.C., X.Y. and H.B. conducted the phylogenetic analyses. J.L, P.E.A. and M.Z. performed computerized tomography restorations. M.Z., P.E.A., X.Y. and B.C. discussed the results and prepared the manuscript.

Author Information The LSIDs urn:lsid:zoobank.org:pub:1CE67C34-52CC-4CEF-B731-DF1A4E85DE4B (article), urn:lsid:zoobank.org:act:D88DD67F-1159-4E7E-AE4C-814249337351(genus), and urn:lsid:zoobank.org:act:77F76728-FBC1-407D-B2C1-1C5D5BF9B59C (species) have been deposited in ZooBank. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.Z. (zhumin@ivpp.ac.cn).

METHODS

Phylogenetic analysis. To determine the position of *Entelognathus primordialis* gen. et sp. nov., and to evaluate the impact of its characters on gnathostome phylogeny, we conducted phylogenetic analyses using a modified data set comprising a total of 253 characters and 75 taxa (see Supplementary Information for details). This data set (hereafter referred to as the full data set) is based on a previous one[16], with revised codings for 29 of the original 138 characters, and with the addition of 115 characters and 15 taxa, including *Entelognathus*. We also analysed subsets of the full data set to detect the effect of including *Entelognathus* and the effect of revising codings, adding characters and adding taxa.

To test the overall homology hypothesis between placoderm and osteichthyan dermal skeletons properly, we followed the precedent of refs 15 and 16 in mostly using neutral descriptive codings for dermal bone characters that potentially span the placoderm–osteichthyan divide. For example, our character 161 is defined as 'number of marginal bones alongside paired median skull roofing bones over the otico-occipital division of braincase', without specifically addressing the relationship between the osteichthyan (tabular, supratemporal) and placoderm (anterior paranuchal, marginal) nomenclatures for bones in this position. This conservative coding strategy is appropriate for the initial phylogenetic assessment of a putative transitional form like *Entelognathus*.

The character data entry and formatting were performed in Mesquite 2.5 (ref. 36). All characters were treated as unordered, and weighted equally. The data set was subjected to parsimony analysis in PAUP* 4.0b10 (ref. 37). Two agnathan taxa (Galeaspida and Osteostraci) were used as the outgroup. Tree searches were conducted using the heuristic algorithm, with 1,000 (full data set and subset 3) or 10,000 (subsets 1–2) random addition sequence replicates, and with 'maxtrees' set to 'automatically increase'. Bremer decay indices were obtained using command files composed by TreeRot[38] in conjunction with the heuristic search algorithm in PAUP*. Bootstrap values were calculated from 1,000 pseudoreplicates using the heuristic search option in PAUP* (random addition sequence with 10 replicates).

For the full data set, Bayesian inference analysis was conducted with MrBayes 3.1.2 (refs 39, 40). Galeaspida was set as the outgroup, and the codings showing polymorphism were changed to '?'. Priors were kept at their default settings for standard (= morphological) analyses. The analysis was run for 10^7 generations to ensure that the average standard deviation of split frequencies is below 0.01 (an indication for convergence of two runs or stationarity). Samples were taken every 10^2 generations, resulting in a total of 10^5 samples for each of the parallel analyses. The first 2.5×10^4 samples for each run, representing the 'burn-in' period, were discarded. The 50% majority-rule consensus tree was computed for the sampled generations.

X-ray micro-computerized tomography. The scanning was carried out using the 225 kV micro-computerized tomography (developed by the Institute of High Energy Physics, Chinese Academy of Sciences (CAS)) at the Key Laboratory of Vertebrate Evolution and Human Origins, CAS. The specimen was scanned with beam energy of 130 kV and a flux of 90 μA at a detector resolution of 10.8 μm per pixel using a 720° rotation with a step size of 0.5° and an unfiltered aluminium reflection target. A total of 1,440 transmission images were reconstructed in a $2{,}048 \times 2{,}048$ matrix of 1,563 slices using a two-dimensional reconstruction software developed by the Institute of High Energy Physics, CAS. The three-dimensional reconstructions were created in Mimics (version 14.12), and images of the reconstructions were exported from Mimics and finalized in Adobe Photoshop and Adobe Illustrator.

36. Maddison, W. P. & Maddison, D. R. *Mesquite: A Modular System For Evolutionary Analysis* v. 2.5 (http://mesquiteproject.org, 2008).
37. Swofford, D. L. *PAUP*: Phylogenetic analysis using parsimony (* and other methods), version 4.0b 10* (Sinauer Associates, 2003).
38. Sorenson, M. D. *TreeRot. Program and documentation* v. 2 (Boston Univ., 1999).
39. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19,** 1572–1574 (2003).
40. Huelsenbeck, J. P. & Ronquist, F. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17,** 754–755 (2001).

doi:10.1038/nature12639

Functional interaction between autophagy and ciliogenesis

Olatz Pampliega[1], Idil Orhon[2,3], Bindi Patel[1], Sunandini Sridhar[1], Antonio Díaz-Carretero[1], Isabelle Beau[3], Patrice Codogno[2,3], Birgit H. Satir[4], Peter Satir[4] & Ana Maria Cuervo[1,4,5]

Nutrient deprivation is a stimulus shared by both autophagy and the formation of primary cilia. The recently discovered role of primary cilia in nutrient sensing and signalling motivated us to explore the possible functional interactions between this signalling hub and autophagy. Here we show that part of the molecular machinery involved in ciliogenesis also participates in the early steps of the autophagic process. Signalling from the cilia, such as that from the Hedgehog pathway, induces autophagy by acting directly on essential autophagy-related proteins strategically located in the base of the cilium by ciliary trafficking proteins. Whereas abrogation of ciliogenesis partially inhibits autophagy, blockage of autophagy enhances primary cilia growth and cilia-associated signalling during normal nutritional conditions. We propose that basal autophagy regulates ciliary growth through the degradation of proteins required for intraflagellar transport. Compromised ability to activate the autophagic response may underlie some common ciliopathies.

Autophagy is a cellular catabolic process that contributes to quality control and maintenance of the cellular energetic balance through the turnover of proteins and organelles in lysosomes[1]. Induction of autophagy recruits proteins and lipids from different intracellular membranes[2] to initiate the formation of autophagosomes, double-membrane vesicles that sequester cytoplasmic material and deliver it through vesicular fusion to lysosomes for degradation[3].

The primary cilium is a non-motile signalling organelle that grows in a specific region of the plasma membrane and senses, among other things, changes in the enrichment of nutrients in the environment[4]. Cargo trafficking along the ciliary axoneme (intraflagellar transport; IFT) is maintained through motor proteins (kinesins and dyneins) and two large multiprotein complexes (IFT particles A and B)[5]. Some subunits of these complexes can be found in other cellular compartments such as Golgi, from where they facilitate mobilization of specific cargo to the basal body and cilium for ciliogenesis and ciliary signalling[6]. The primary cilium coordinates a variety of signalling pathways including the Hedgehog (Hh) pathway, which requires IFT-mediated recruitment to the basal body and axoneme of smoothened (SMO) and the transcription factors GLI1 and GLI2[7,8].

In many types of cultured cells, deprivation of serum from the media induces almost linear growth of primary cilia for up to 2 days[9]. Induction of autophagy also occurs in the first hours that follow serum removal, and can be sustained for the full duration of the starvation period[10]. Despite the temporal coincidence of the formation of autophagosomes and primary cilia, any possible relation between the autophagic and ciliogenesis machineries remains unknown. Likewise, the possible involvement of autophagy in ciliary dynamics has not been analysed.

In this study we investigate the functional interaction between primary cilia and autophagic induction. We have found that maximum activation of autophagy in response to nutrient deprivation requires the strategic location of components of the autophagic machinery at the ciliary base in a IFT and Hh signalling-dependent manner. In contrast, when autophagy is compromised, ciliogenesis is enhanced and cilia grow longer. We conclude that ciliary signalling pathways, such as Hh, may specify a cilia-mediated autophagy, closely related to autophagosome assembly at the plasma membrane and that, in turn, basal autophagy negatively modulates ciliary growth through turnover of essential ciliogenesis proteins (Extended Data Fig. 1).

Disruption of IFT compromises autophagy

To determine the role of primary cilia on autophagy, we used two cellular models with compromised ciliogenesis: mouse embryonic fibroblasts (MEFs) stably knocked down for IFT20 ($IFT20^{-}$) and kidney epithelial cells (KECs) from mice with a hypomorphic mutant IFT88 (*Ift88*$^{-/-}$), two IFT-B complex components required for ciliogenesis[6] (Extended Data Fig. 2a). After 24 h of serum removal, cells with detectable primary cilia were reduced by 60% in $IFT20^{-}$ MEFs and almost absent in the *Ift88*$^{-/-}$ KECs (Fig. 1a).

Rates of lysosomal protein degradation after serum deprivation, a stimulus that activates autophagy, were reduced in the two ciliogenesis-impaired models (Extended Data Fig. 2b). To test whether autophagy was defective we analysed flux through the autophagic pathway using lysosomal proteolysis inhibitors and measured changes in levels of LC3-II, an autophagosome component degraded with the cargo in lysosomes[11]. The characteristic upregulation of the autophagic flux after serum removal was reduced in $IFT20^{-}$ MEFs whereas basal autophagic flux (rich media) was only discretely reduced (Fig. 1b). Although a high percentage of KECs are ciliated under basal conditions and serum removal has small effect on ciliogenesis (Extended Data Fig. 2c), we still observed a reduction in autophagic flux upon serum removal in *Ift88*$^{-/-}$ KECs when compared to wild type (Fig. 1b), indicating that primary cilia are not necessary to maintain the high basal autophagic flux of these cells (Extended Data Fig. 2d), but are essential to sustain proper autophagic flux during nutrient deprivation.

To directly analyse autophagosome formation and their clearance by lysosomes, we transfected cells with a pH-sensitive reporter (mCherry–GFP–LC3; fusion of mCherry, green fluorescent protein and LC3) that highlights autophagosomes as yellow puncta and

[1]Department of Development and Molecular Biology, Albert Einstein College of Medicine, Bronx, New York 10461, USA. [2]INSERM U845; Paris-Descartes University, 75014 Paris, France. [3]INSERM U984; University Paris-Sud 11; 92296 Châtenay-Malabry, France. [4]Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, Bronx, New York 10461, USA. [5]Institute for Aging Studies, Albert Einstein College of Medicine, Bronx, New York 10461, USA.

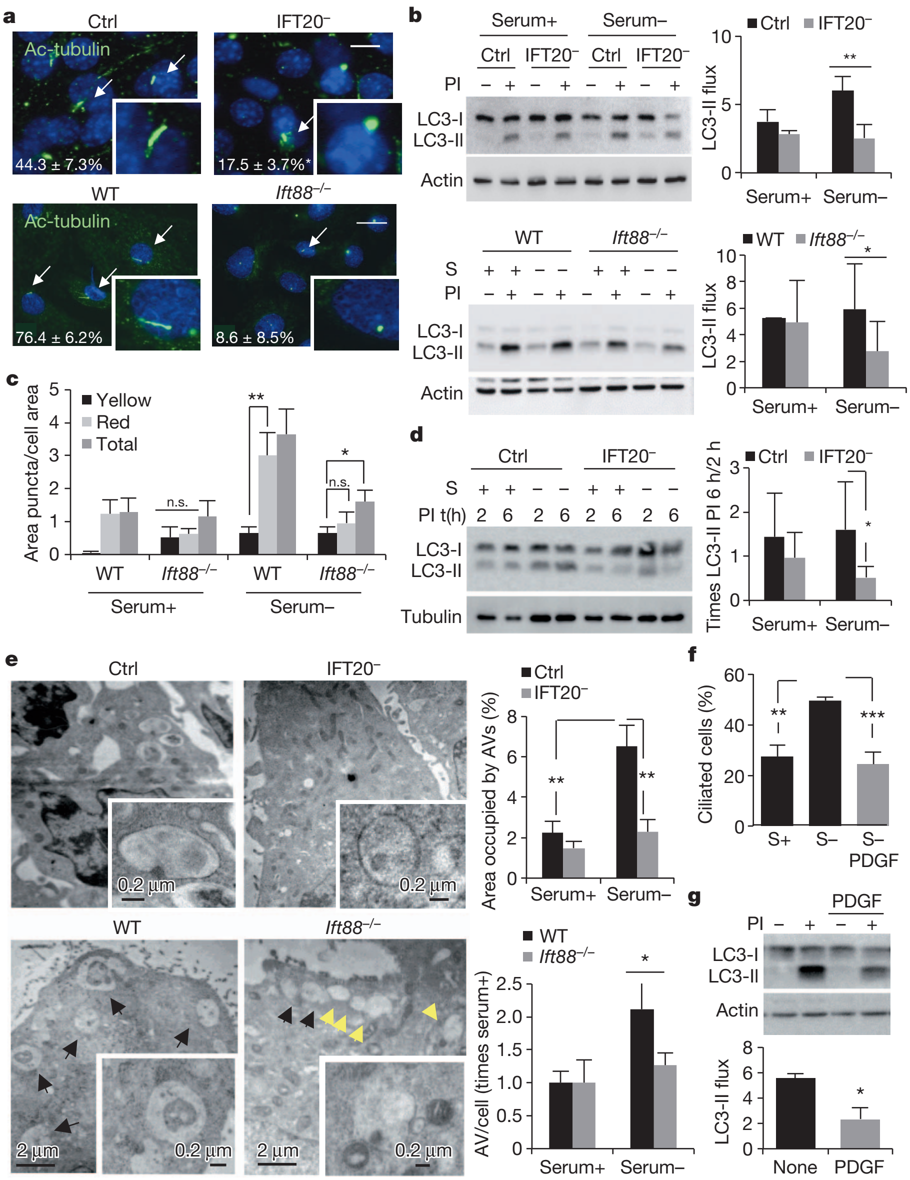


Figure 1 | Blockage of IFT reduces autophagic activity. **a**, Percentage of ciliated cells in control (Ctrl) and IFT20^{-} MEFs (*$P = 0.030$, $n = 3$) and wild-type (WT) and *Ift88*$^{-/-}$ KECs (**$P = 0.003$, 25 cells each, $n = 3$) after 24 h of serum removal. Arrows, cilia. **b**, LC3 immunoblot (left) for autophagic flux quantification (right) in Ctrl and IFT20^{-} MEFs (top, **$P = 0.0003$, $n = 7$), and in WT and *Ift88*$^{-/-}$ KECs (bottom, *$P = 0.035$, $n = 9$). PI, protease inhibitors; S, serum. **c**, Autophagic flux in WT and *Ift88*$^{-/-}$KECs by mCherry–GFP–LC3 puncta quantification. Puncta: yellow, autophagosomes; red only, autophagolysosomes; total, both together (**$P = 0.008$, *$P = 0.025$). **d**, LC3 immunoblot (left) for quantification of autophagosome formation (right). (*$P = 0.048$, $n = 4$). **e**, Electron microscopy images (left) and morphometric quantification of autophagic vacuoles (AV, right) in Ctrl and IFT20^{-} MEFs (top, **$P = 0.001$, 10 fields in 2 experiments) and KECs WT and *Ift88*$^{-/-}$ (bottom, *$P = 0.037$, 5 fields). Arrows, autophagic vacuoles (AV, black) and endosomal compartments (yellow). **f**, Percentage of ciliated MEFs after induction of cilia resorption with 50 ng ml^{-1} platelet-derived growth factor (PDGF) treatment. (**$P = 0.001$, ***$P = 0.0009$, 121, 230 and 162 cells, $n = 5$). **g**, LC3 immunoblot (top) for autophagic flux quantification (bottom) upon PDGF treatment (*$P =$ 0.005, $n = 3$). Scale bars in **a**, 10 μm. n.s., statistically non-significant. Mean ± s.d. in **b**, **d** and mean ± s.e.m. in other panels.

autophagolysosomes (post-lysosomal fusion) as red puncta[12]. Basal levels of autophagic vacuoles were comparable in wild-type and *Ift88*$^{-/-}$ KECs, but upon removal of serum *Ift88*$^{-/-}$ KECs showed a significantly lower content of autophagolysosomes (Fig. 1c and Extended Data Fig. 2e). This reduction in autophagolysosomes was not paralleled by an increase in autophagosomes, and the overall content of LC3-positive vesicles in serum-deprived *Ift88*$^{-/-}$ KECs was lower than in wild-type cells (Extended Data Fig. 2f), indicating that reduced autophagic flux was mainly owing to lower autophagosome synthesis rather than blockage in their degradation. Autophagosome formation (assessed as the increase in LC3-II levels at two time points after inhibition of lysosomal proteolysis) was also reduced in IFT20^{-} MEFs (Fig. 1d). Ultrastructural analysis confirmed that despite normal autophagosome morphology, IFT20^{-} and *Ift88*$^{-/-}$ cells failed to expand the autophagic compartment in response to serum deprivation (Fig. 1e).

Contrary to the restricted location of IFT88 to the ciliary base[13], IFT20 is also present in other cellular compartments[6]. However, the effect of IFT20 on autophagy seems to be primarily related to its function in ciliogenesis because additional knockdown for IFT20 in IFT88 knockdown cells did not result in further reduction of autophagy (Extended Data Fig. 3a–d), supporting that blockage of anterograde IFT (aIFT) compromises activation of inducible autophagy. Consistent with this notion, cilia resorption induced by treatment with platelet-derived growth factor[14] (Fig. 1f) markedly decreased starvation-induced autophagic flux in MEFs, KECs and in RGC-5 cells, a retinal ganglion cell line with robust induction of ciliogenesis in response to serum removal (Fig. 1g and Extended Data Fig. 3e).

Overall these findings support a reduced ability for autophagosome synthesis in cells with compromised ciliogenesis that limits their capacity to upregulate autophagy in response to nutrient deficiency.

Ciliary Hh signalling induces autophagy

Altered aIFT blocks ciliogenesis and prevents the recruitment of essential signalling molecules to the ciliary region[6]. We proposed that malfunctioning of ciliary signalling may be behind the defective starvation-induced autophagy in cells with compromised IFT. Although primary cilia downregulate signalling through the mammalian target of rapamycin (mTOR) pathway[15], a well-known negative regulator of autophagy, treatment of IFT20^{-} and *Ift88*$^{-/-}$ cells with the mTOR inhibitor

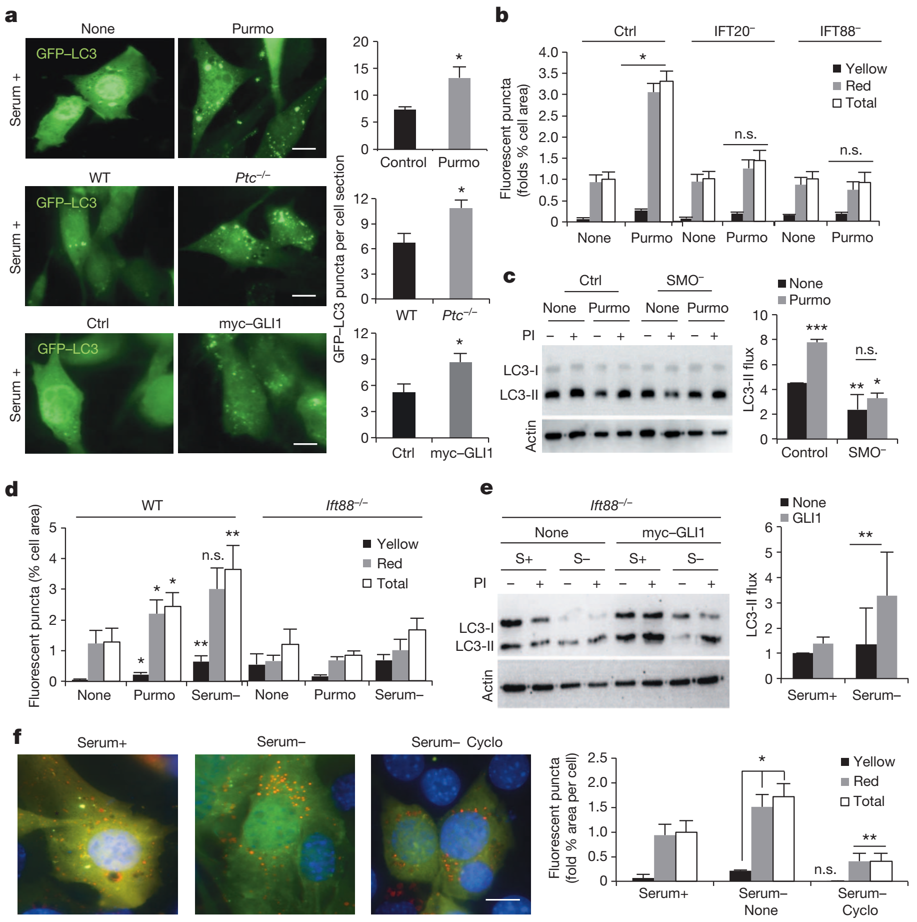


Figure 2 | Hedgehog signalling requires the primary cilia to regulate autophagy. **a**, GFP–LC3 puncta pattern (left) and quantification (right) in MEFs treated with purmorphamine (Purmo; top, $*P = 0.028$, $n = 4$), in *Ptc*$^{-/-}$ MEFs (middle, $*P = 0.015$, $n = 4$) and after myc–GLI1 overexpression (bottom, $*P = 0.012$, $n = 3$). **b**, Autophagic flux by mCherry–GFP–LC3 puncta quantification upon purmorphamine treatment in MEFs in the presence of serum. Yellow, autophagosomes; red only, autophagolysosomes; total, both compartments. ($*P = 0.0323$, 40 fields). **c**, LC3 immunoblot (left) for autophagic flux quantification (right) in Ctrl ($**P = 0.001$) and MEFs knocked down for SMO (SMO−) ($**P = 0.0001$) upon purmorphamine treatment ($*P = 0.009$, $n = 4$). **d**, Autophagic flux by mCherry–GFP–LC3 puncta quantification in WT and *Ift88*$^{-/-}$ KECs upon purmorphamine treatment ($*P = 0.046$, $**P = 0.008$, 40 fields). **e**, LC3 immunoblot (left) for autophagic flux quantification (right) in *Ift88*$^{-/-}$ KECs overexpressing myc–GLI1 ($**P = 0.00006$, $n = 3$). **f**, Autophagic flux by mCherry–GFP–LC3 puncta quantification in MEFs treated with cyclopamine upon serum removal ($*P = 0.011$, $**P = 0.013$, 25 fields). Scale bars, 10 μm. Mean ± s.d. in **c**, **f** and mean ± s.e.m. in other panels.

rapamycin failed to restore normal autophagic activity in these cells (data not shown); thereby reduced autophagy in ciliogenesis-defective cells was not due to enhanced mTOR signalling.

We next focused on the Hedgehog (Hh) signalling pathway because of its dependence on the primary cilium and intact IFT[7,8]. Treatment of serum-supplemented MEFs or KECs with purmorphamine, an SMO agonist that activates expression of Hh downstream factors (Extended Data Fig. 4a, c), returned content of LC3-positive compartments and autophagic flux to values observed upon serum removal (Fig. 2a–d and Extended Data Fig. 4d–f). Purmorphamine failed to induce autophagy in IFT20^{-} or IFT88^{-} MEFs and in *Ift88*$^{-/-}$ KECs, supporting dependence on IFT (Fig. 2b, d and Extended Data Fig. 4f) and in SMO^{-} cells (Fig. 2c), which as expected showed reduced purmorphamine-mediated upregulation of GLI1 and GLI2 (Extended Data Fig. 4c). Two other conditions that activate Hh signalling, Patched-1 receptor knockout (*Ptc*$^{-/-}$, also known as *Ptch1*; constitutive activation of Hh signalling and ciliogenesis[8]) and overexpression of GLI1 also upregulated autophagy (Fig. 2a and Extended Data Fig. 4d, g–j), and in fact, overexpression of GLI1 is sufficient to partially rescue the autophagic defect in *Ift88*$^{-/-}$ cells (Fig. 2e). Conversely, two interventions that reduce Hh signalling, knockdown of SMO and treatment with Hh antagonist cyclopamine, reduced starvation-induced autophagy (Fig. 2c, f and Extended Data Fig. 4k).

Altogether, these data reveal a positive regulatory effect of Hh signalling on autophagy, and support that the inability to activate autophagy in cells with defective IFT originates, at least in part, from the loss of Hh signalling.

Autophagic machinery localizes at the cilia

Besides ciliogenesis, IFT also participates in recruitment of non-ciliary proteins to the plasma membrane[16], a site of autophagosome formation[17], which made us propose an involvement of aIFT in the delivery of the autophagic machinery to this membrane.

Co-immunostaining for different autophagy-related proteins (ATGs, green) and acetylated tubulin (red) to highlight the ciliary axoneme in serum-deprived KECs revealed that of the twelve ATGs analysed, five ATGs[2,3] (ATG16L, AMBRA1, LC3, GABARAP and VPS15) localized as discrete puncta along the ciliary axoneme, clearly distinguishable from the surrounding cytosol signal with deconvolution immunofluorescence and three-dimensional (3D) wire modelling (Fig. 3a, Extended Data Fig. 5a, b and Supplementary Videos 1–5). Similar colocalization was observed using fluorescent-tagged versions of the ciliary protein inversin[18] or of ATGs (Extended Data Fig. 5c, d). Other more abundant ATGs, such as ATG14, were not detectable in the axoneme, and knockdown against cilia-associate ATGs abolished their staining (Extended Data Fig. 5b, d) supporting specificity of the staining. These ATGs were also detectable in primary cilia isolated by the 'peeling off' procedure (Fig. 3b).

A larger number of ATGs associated with the basal body (highlighted with anti-gamma tubulin). In addition to the five cilia-associated ATGs, we also found ATG14, VPS34, ATG7 and ATG5, but not Beclin 1 or ULK-1, at the base of the axoneme (Fig. 3c; LAMP-2 is used a negative control). Similar colocalization experiments in serum-supplemented cells, to repress autophagy, and in cells with compromised ciliogenesis (*Ift88*$^{-/-}$ cells) revealed three types of association of ATGs with the basal body: serum- and aIFT-dependent, such as VPS34 and ATG16L; serum-independent but aIFT-dependent, as ATG7 and ATG14; and serum- and aIFT-independent, as ATG5 and LC3 (Fig. 3d–f and Extended Data Fig. 6).

Disruption of IFT also altered the overall intracellular distribution of ATGs that localize at the basal body in an aIFT-dependent manner. Thus, ATG7 was normally observed in KECs as discrete cytosolic puncta

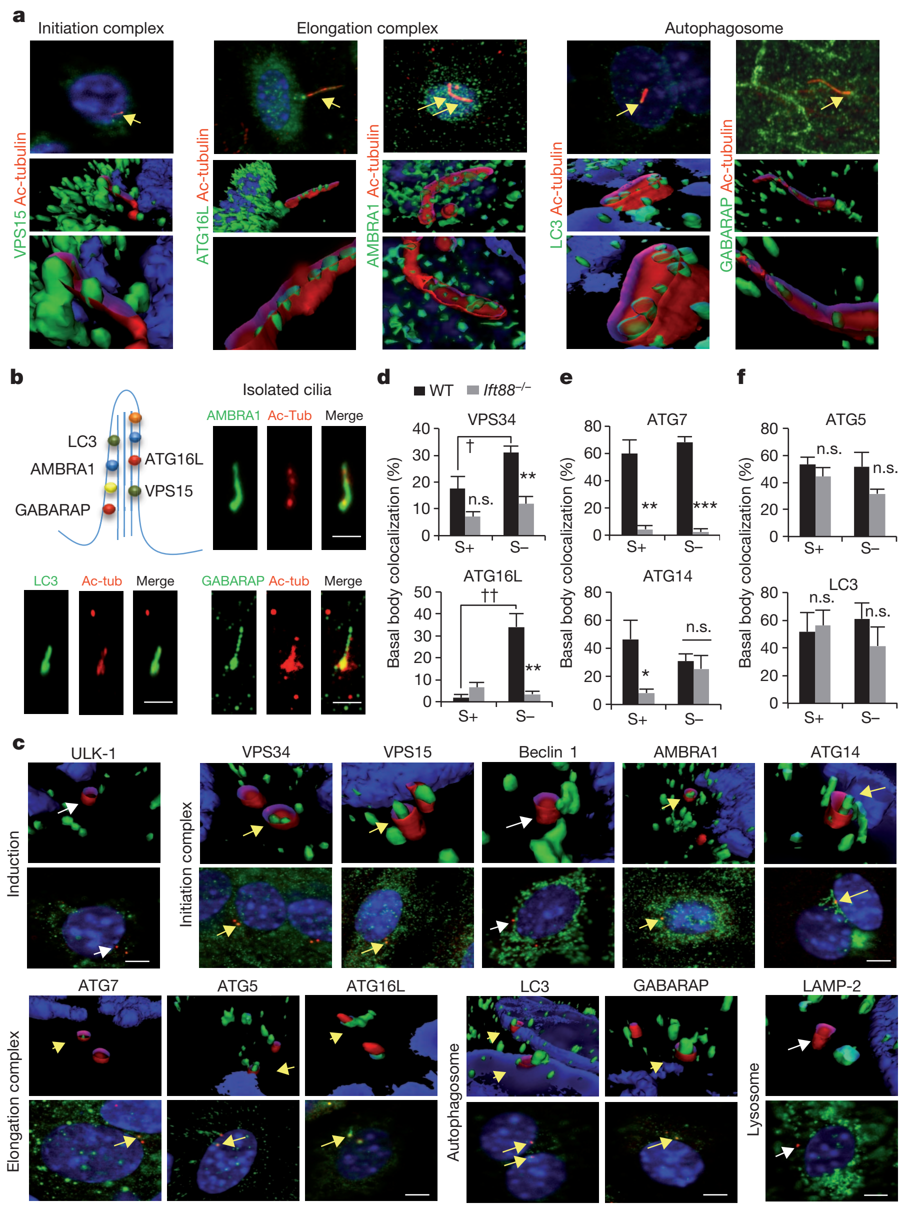


Figure 3 | Autophagy-related proteins associate with ciliary structures in a serum-dependent manner. **a**, Co-immunostaining (top) and 3D reconstruction (middle and bottom) for the indicated autophagy-related proteins (ATGs, green) and acetylated tubulin (red, ciliary axoneme marker) in KECs maintained in the absence of serum for 24 h. Yellow arrows, colocalization. **b**, Schematic representation of the ATGs associated with the axoneme (top left). Co-immunostaining for ATGs and acetylated tubulin in cilia isolated by peeling off. Individual channels and merged images are shown. **c**, Co-immunostaining (bottom) and 3D reconstruction (top) for ATGs (green) and gamma-tubulin (red, basal body marker) in the same cells and conditions. Arrows, colocalization (yellow), no colocalization (white). **d–f**, Percentage of cells with colocalizing ATGs in the basal body (BB) in WT and *Ift88*$^{-/-}$ KECs in the presence or absence of serum. **d**, Serum- and IFT-dependent association of ATGs with the BB (VPS34, †P = 0.04, **P = 0.002, n = 4; ATG16L, ††P = 0.0003, **P = 0.0009, 15 cells each per experiment, n = 7). **e**, IFT-dependent but serum-independent association of ATGs with the BB (**P = 0.001, ***$P = 9.33768 \times 10^{-6}$, n = 4; *P = 0.018, 15 cells each per experiment, n = 5). **f**, BB association of ATGs independent on serum or IFT (ATG5, n = 5; LC3, n = 8; 15 cells each per experiment). Scale bars, 10 μm. Mean ± s.e.m.

that markedly increased upon starvation, but it adopted a diffuse reticular pattern unchanged by starvation in *Ift88*$^{-/-}$ KECs (Extended Data Fig. 7a). Disruption of IFT also reduced the basal association of ATG14, key for autophagy initiation[3], with the Golgi and its mobilization into discrete cytosolic puncta upon serum removal (Extended Data Fig. 7a).

Interestingly, IFT88 deficiency also blunted the starvation-induced changes in the intracellular location of ATGs that localize to the basal body independently of IFT88, such as LC3 and ATG5 (Fig. 3f). The association with the plasma membrane of LC3 and GABARAP (both involved in autophagosome membrane formation/elongation[3]), clearly noticeable in wild-type cells upon serum removal, was no longer observed in the IFT88-deficient cells (Extended Data Fig. 7a). Likewise, the often ‘hook-like’ organization of ATG5 (involved in autophagosome membrane elongation) at the basal body was also abrogated in *Ift88*$^{-/-}$ KECs (Extended Data Fig. 7b, c). The abundance of ATG5 hook-like structures under basal conditions in cells with constitutive Hh signalling (*Ptc*$^{-/-}$; Extended Data Fig. 7d) indicates the need of a functional cilium for ATG5 clustering in this location.

These findings support a role for aIFT in the cellular relocation of the autophagic machinery to sites of autophagosome formation.

IFT-dependent trafficking of ATG16L

To gain a better understanding of the mechanisms behind the ciliary regulation of starvation-induced autophagy, we focused on ATGs that associated preferentially with ciliary structures upon serum removal. We focused on ATG16L because it was almost undetectable in the basal body in serum-supplemented cells (Fig. 3d), starvation induced its IFT-dependent association with the basal body and its presence in this location increased in two models with enhanced ciliary Hh signalling (Extended Data Fig. 8a, b). We proposed that active recruitment of ATG16L to the ciliary base during starvation could be the trigger for ciliary-induced autophagy.

We first investigated the contribution of IFT20 to ciliary delivery of ATG16L because IFT20 also participates in trafficking of ciliary membrane proteins from the Golgi to the base of the cilium[6]. To analyse separately both functions of IFT20, we compared its association with ATG16L in cells with intact or disrupted aIFT (*Ift88*$^{-/-}$ KECs). We found that IFT20 and ATG16L colocalized in small cytosolic vesicles that become more abundant upon serum removal and that this colocalization was only partially reduced when aIFT was disrupted (Extended Data Fig. 8c). Both proteins co-immunoprecipitated independently of

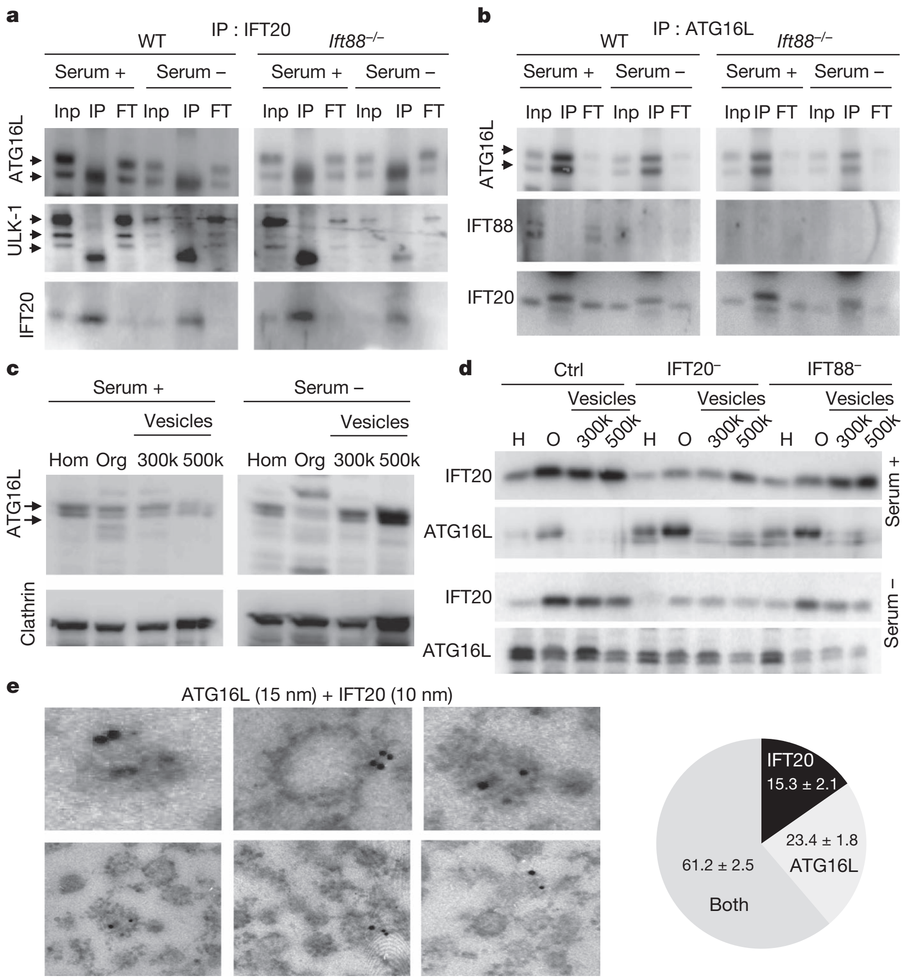


Figure 4 | IFT20 regulates trafficking of ATG16L to the cilium. **a**, **b**, Immunoblot for the indicated proteins of WT and *Ift88*$^{-/-}$ KECs subjected to co-immunoprecipitation of IFT20 (**a**) and ATG16L (**b**). Inp, 1/10 input; IP, immunoprecipitate; FT, 1/10 flow-through. **c**, Immunoblot for ATG16L and clathrin in MEF homogenate (Hom), and the pellets from 1 h centrifugation at 100,000*g* (organelles (Org)), 300,000*g* (300k) and 500,000*g* (vesicles, 500k). **d**, Immunoblot for IFT20 and ATG16L in the same fractions isolated from Ctrl, IFT20$^-$ and IFT88$^-$ MEFs in the presence (top) and absence (bottom) of serum. **e**, Immunogold electron microscopy for ATG16L (15-nm particles) and IFT20 (10-nm particles) in vesicles from MEFs isolated by the same method as in **c**. Right, quantification of IFT20 and ATG16L presence in the vesicles (11 fields). Mean ± s.e.m.

the presence of IFT88 (Fig. 4a), and IFT20, but not IFT88, could be pulled-down with ATG16L (Fig. 4b). Starvation did not change the association between ATG16L and IFT20, but changed the location of the IFT20- and ATG16L-positive vesicles from the proximity of the Golgi towards a more cytosolic location (Extended Data Fig. 8c), indicating that the same signal that enhances IFT20-mediated trafficking of proteins from the Golgi to the cilia may also be responsible for ciliary delivery of ATG16L. Other ATGs (VPS15, ATG7) have minimal colocalization with IFT20, or colocalize only at the Golgi (ATG14) (Extended Data Fig. 8d), and none of them binds directly to IFT20 (Extended Data Fig. 8e–g).

The amount of ATG16L in isolated clathrin-enriched vesicles[19] was very low under basal conditions, but it increased markedly upon serum removal (Fig. 4c). Although the vesicular content of IFT20 did not change with starvation (Fig. 4d), knockdown of IFT20 reduced starvation-induced loading of ATG16L in the vesicles (Fig. 4d). Interestingly, ATG16L content was also reduced in vesicles isolated from serum-deprived cells knocked down for IFT88 (Fig. 4d), explaining the reduced colocalization between IFT20 and ATG16L in *Ift88*$^{-/-}$ cells. Electron microscopy and double immunogold labelling for IFT20 and ATG16L confirmed the vesicle nature of the fractions and the coincidence of both proteins in more than 60% of the labelled vesicles (Fig. 4e and Extended Data Fig. 8h).

We propose that, during starvation, ATG16L reaches the base of the cilia through the Golgi-to-cilia shuttling function of IFT20, and then it accesses the ciliary region by IFT88-dependent mechanisms[20]. Overall, our results support a novel role for the centrioles as organization centres for the autophagic machinery and for IFT proteins as common components shared by both ciliogenesis and starvation-induced autophagosome formation.

Autophagy activation reduces cilia growth

Although inducible autophagy relies on IFT and cilia function, we found that autophagy is not required for cilia formation. In fact, *Atg5*$^{-/-}$ MEFs formed cilia both longer and faster than wild-type MEFs upon serum removal, and a higher percentage of them grew cilia even in basal conditions (Fig. 5a–d). Knockdown of two other ATGs and even chemical inhibition of autophagy were sufficient to induce ciliogenesis in different cell types under basal conditions, whereas upregulation of autophagic activity with rapamycin did not affect either basal or inducible ciliary growth (Extended Data Fig. 9).

Scanning electron microscopy revealed that in *Atg5*$^{-/-}$ MEFs starvation induced formation of long and narrow cilia that come out of smaller ciliary pockets, and confirmed that *Atg5*$^{-/-}$ MEFs grow primary cilia in serum-supplemented media, although shorter and with fewer surface-adhered vesicles (presumably exosomes) than the cilia grown during starvation (Fig. 5c and Extended Data Fig. 10). Activation of ciliary Hh signalling by starvation or treatment with purmorphamine showed efficient recruitment of SMO to the cilia in *Atg5*$^{-/-}$ cells and increased expression of the downstream transcription factors GLI1 and GLI2 in these cells even under basal conditions, supporting full functionality of the cilia induced by autophagic blockage (Fig. 5e, f and Extended Data Fig. 11a).

To start investigating the basis for this inhibitory effect of autophagy on ciliogenesis, we compared levels and distribution of IFT proteins in wild-type and *Atg5*$^{-/-}$ MEFs and found that basal levels of IFT20, but not IFT88, were markedly higher in *Atg5*$^{-/-}$ MEFs (Fig. 5g). Because chemical inhibition of lysosomal proteolysis in control cells reproduces a similar increase in IFT20 levels (Fig. 5h), and IFT20 could be detected in isolated autophagic vacuoles (Fig. 5i), we propose that basal levels of IFT20, and consequently ciliary

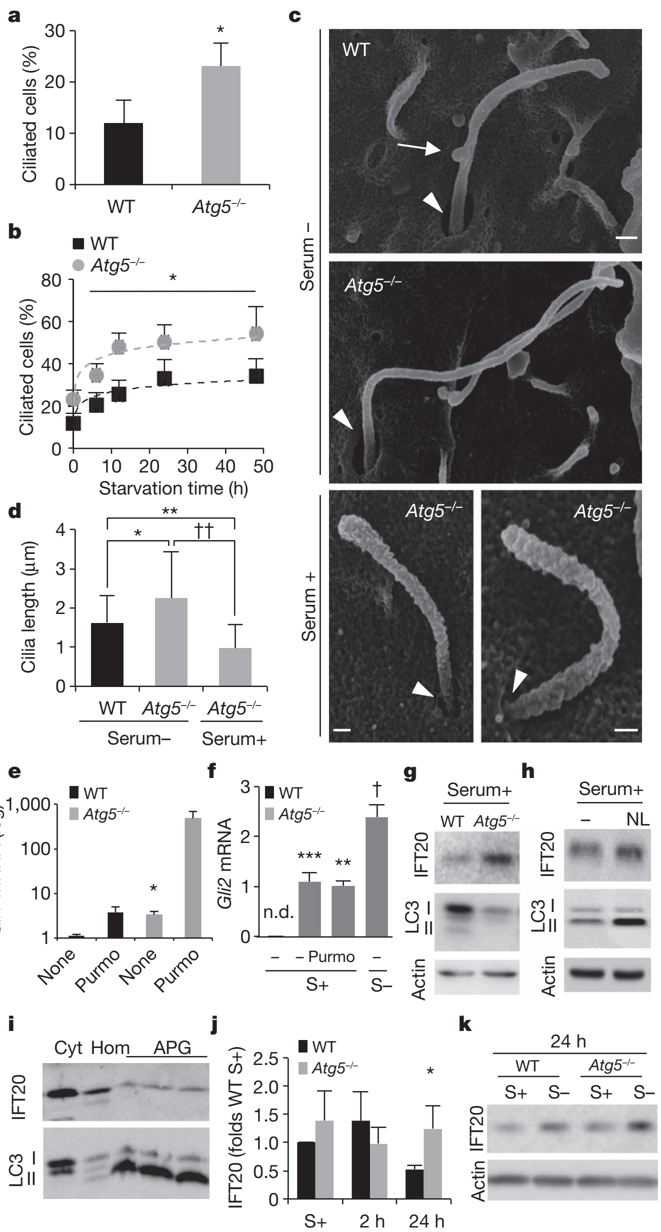


Figure 5 | Ciliogenesis is enhanced in autophagy-defective cells. **a**, **b**, Ciliated *Atg5*$^{-/-}$ MEFs in serum^{+} (*$P = 0.038$, 25 cells each per experiment, $n = 3$) (**a**) or serum^{-} (*$P = 0.006$, nonlinear fit regression, 25 cells each per experiment, $n = 3$) (**b**). **c**, Scanning electron microscopy (SEM) of primary cilia. Arrows, cilia-associated vesicles; arrowheads, ciliary pocket. **d**, Cilia length quantification from SEM (*$P = 0.027$ $n = 28$; **$P = 0.001$, ††$P = 0.0001$, $n = 23$). **e**, **f**, GLI1 (*$P = 0.04$, $n = 3$) (**e**) and GLI2 (***$P = 0.0005$, **$P = 0.0046$, †$P = 0.0132$, $n = 3$) (**f**) messenger RNA expression in *Atg5*$^{-/-}$ MEFs. **g**, **h**, IFT20 immunoblot in *Atg5*$^{-/-}$ (**g**) and WT MEFs with lysosomal inhibitors (NL) (**h**). **i**, IFT20 immunoblot in cytosol (Cyt), homogenate (Hom) and autophagosomes (APG). **j**, IFT20 protein levels in *Atg5*$^{-/-}$ MEFs (*$P = 0.038$, $n = 6$). **k**, IFT20 immunoblot in *Atg5*$^{-/-}$ MEFs. n.s., statistically non-significant. Mean ± s.d. in **d** and **j** and mean ± s.e.m. in other panels.

growth, are regulated at least in part through IFT20 degradation by basal autophagy. Although most IFT20 associates with Golgi, the basal increase in cellular IFT20 levels in *Atg5*$^{-/-}$ MEFs seems to occur in the small cytosolic vesicles only visible in control cells during starvation-induced ciliogenesis (Extended Data Fig. 11b). Interestingly, the differences in the IFT20 content between wild-type and *Atg5*$^{-/-}$ MEFs were no longer observed early after serum removal, but become apparent again after 24 h in serum-deprived media (Fig. 5j, k). Studies with lysosomal proteolysis inhibitors in wild-type MEFs confirmed absence of lysosomal degradation of IFT20 during the first 12 h of serum removal, when inducible autophagy is maximally activated (Extended Data Fig. 11c, d). In light of these findings, we propose that basal autophagy contributes to the regulation of ciliogenesis and the length of an already formed cilium, at least in part through the degradation of IFT20. The switch from basal to induced autophagy seems to spare cytosolic IFT20 from degradation and allow its active engagement in the vesicular trafficking required for ciliogenesis.

Discussion

In this work, we identify a previously unknown reciprocal relationship between primary cilium and autophagy whereby components of the ciliary machinery are shared with autophagy for starvation-induced autophagosome biogenesis. In contrast, basal autophagy inhibits ciliogenesis by limiting trafficking to the cilium of components required for ciliary growth. The temporal coincidence of activation of ciliogenesis and autophagy at the early times of nutritional deprivation may thus serve as a self-regulatory brake for each of these processes (Extended Data Fig. 1).

The localization of autophagy initiating ATGs at the ciliary base and the active recruitment of ATG16L to this location upon starvation suggests that sensing of nutrient deficiency and activation of signalling from the cilium could initiate a cilia-mediated autophagic program (see additional Discussion 1 in Supplementary Information). The presence along the cilium axoneme of the pre-autophagosomal ATG16L and integral autophagosome membrane ATGs—for example, LC3 and GABARAP—and their IFT-dependent enrichment at the plasma membrane suggest that cilia-mediated autophagy may induce autophagosome formation from this location, making the ciliary pocket, characterized by high vesicular activity[21], an attractive place for autophagosome formation (see additional Discussion 2 in Supplementary Information).

The negative regulatory role of basal autophagy on ciliogenesis identified here may take place through different mechanisms. We have found that basal autophagy controls trafficking of ciliary proteins by limiting the amount of IFT20 accessible for shuttling between the Golgi and the ciliary base (see additional Discussion 3 in Supplementary Information). As induction of autophagy depends on the same IFT20 protein that autophagy degrades, this could represent a novel mechanism for self-containment of the autophagic process.

The primary cilium is constitutively present in most tissues, but the fact that ciliogenesis blockage in cells with permanent cilia compromises their starvation-induced autophagy supports that it is not the mere presence of a cilium but the activation of ciliary signalling by starvation that contributes to autophagic induction. We have also identified the regulatory effect of the ciliary Hedgehog signalling pathway, for which an emerging role in autophagy regulation has been proposed[22–25] (see additional Discussion 4 in Supplementary Information). The possible participation of the cilium in the integration of other autophagy-inducing signals and the effect of autophagy in IFT20 functions beyond ciliogenesis[16] require further investigation.

METHODS SUMMARY

MEFs defective in autophagy (*Atg5*$^{-/-}$) were from N. Mizushima, *Ift88*$^{-/-}$ (also known as *Tg737*$^{-/-}$) mouse kidney epithelial cells were from G. Pazour, and the RGC-5 cell line from P. Boya. Cell culture, removal of serum and treatment with chemical modulators of autophagy were performed as described before[26]. Intracellular protein degradation was measured as described before[27]. Autophagic flux was measured as changes in levels of LC3-II upon inhibition of lysosomal proteolysis[11]. Lentivirus-mediated short hairpin RNA silencing was performed as described before[26]. Cilia were isolated by the peel-off technique[28] and autophagosomes and autophagolysosomes by floatation in metrizamide gradients as previously described[29]. Cells grown on coverslips were processed for immunofluorescence following standard procedures[26] and imaged in an Apotome.2 system using an Axiovert 200 fluorescence microscope (Carl Zeiss). Transmission electron microscopy was performed as before[26] and scanning electron microscopy following standard

procedures[30]. Gel densitometry, morphometric analysis and quantification of colocalization was done using Image J (NIH). Protein quantification, co-immunoprecipitation, electrophoresis and immunoblot and semiquantitative real-time PCR were performed as previously described[26]. Immunoblot membranes were developed using the LAS-3000 Imaging System (Fujifilm). Results are shown as mean ± s.e.m. or mean ± s.d., and Student's t-test for unpaired data was used for statistical analysis and one-way analysis of variance (ANOVA) was used for multiple comparisons. A value of $P < 0.05$ was considered statistically significant.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 August 2012; accepted 10 September 2013.
Published online 2 October 2013.

1. Mizushima, N., Levine, B., Cuervo, A. M. & Klionsky, D. J. Autophagy fights disease through cellular self-digestion. *Nature* **451,** 1069–1075 (2008).
2. Hamasaki, M., Shibutani, S. T. & Yoshimori, T. Up-to-date membrane biogenesis in the autophagosome formation. *Curr. Opin. Cell Biol.* **25,** 455–460 (2013).
3. Yang, Z. & Klionsky, D. J. Mammalian autophagy: core molecular machinery and signaling regulation. *Curr. Opin. Cell Biol.* **22,** 124–131 (2010).
4. Satir, P., Pedersen, L. B. & Christensen, S. T. The primary cilium at a glance. *J. Cell Sci.* **123,** 499–503 (2010).
5. Taschner, M., Bhogaraju, S. & Lorentzen, E. Architecture and function of IFT complex proteins in ciliogenesis. *Differentiation* **83,** S12–S22 (2012).
6. Follit, J. A., Xu, F., Keady, B. T. & Pazour, G. J. Characterization of mouse IFT complex B. *Cell Motil. Cytoskeleton* **66,** 457–468 (2009).
7. Liem, K. F. *et al.* The IFT-A complex regulates Shh signaling through cilia structure and membrane protein trafficking. *J. Cell Biol.* **197,** 789–800 (2012).
8. Rohatgi, R., Milenkovic, L. & Scott, M. P. Patched1 regulates Hedgehog signaling at the primary cilium. *Science* **317,** 372–376 (2007).
9. Kiprilov, E. N. *et al.* Human embryonic stem cells in culture possess primary cilia with hedgehog signaling machinery. *J. Cell Biol.* **180,** 897–904 (2008).
10. Onodera, J. & Ohsumi, Y. Autophagy is required for maintenance of amino acid levels and protein synthesis under nitrogen starvation. *J. Biol. Chem.* **280,** 31582–31586 (2005).
11. Tanida, I., Minematsu-Ikeguchi, N., Ueno, T. & Kominami, E. Lysosomal turnover, but not a cellular level, of endogenous LC3 is a marker for autophagy. *Autophagy* **1,** 84–91 (2005).
12. Klionsky, D. J. *et al.* Guidelines for the use and interpretation of assays for monitoring autophagy. *Autophagy* **8,** 445–544 (2012).
13. Pazour, G. J., San Agustin, J. T., Follit, J. A., Rosenbaum, J. L. & Witman, G. B. Polycystin-2 localizes to kidney cilia and the ciliary level is elevated in orpk mice with polycystic kidney disease. *Curr. Biol.* **12,** R378–R380 (2002).
14. Schneider, L. *et al.* PDGFRαα signaling is regulated through the primary cilium in fibroblasts. *Curr. Biol.* **15,** 1861–1866 (2005).
15. Boehlke, C. *et al.* Primary cilia regulate mTORC1 activity and cell size through Lkb1. *Nature Cell Biol.* **12,** 1115–1122 (2010).
16. Finetti, F. *et al.* Intraflagellar transport is required for polarized recycling of the TCR/CD3 complex to the immune synapse. *Nature Cell Biol.* **11,** 1332–1339 (2009).
17. Ravikumar, B., Moreau, K., Jahreiss, L., Puri, C. & Rubinsztein, D. C. Plasma membrane contributes to the formation of pre-autophagosomal structures. *Nature Cell Biol.* **12,** 747–757 (2010).
18. Veland, I. R. *et al.* Inversin/nephrocystin-2 is required for fibroblast polarity and directional cell migration. *PLoS ONE* **8,** e60193 (2013).
19. Sridhar, S. *et al.* The lipid kinase PI4KIIIβ preserves lysosomal identity. *EMBO J.* **32,** 324–339 (2012).
20. Richey, E. A. & Qin, H. Dissecting the sequential assembly and localization of intraflagellar transport particle complex B in *Chlamydomonas*. *PLoS ONE* **7,** e43118 (2012).
21. Satir, P. The new biology of cilia: review and annotation of a symposium. *Dev. Dyn.* **241,** 426–430 (2012).
22. Li, H. *et al.* Sonic hedgehog promotes autophagy of vascular smooth muscle cells. *Am. J. Physiol. Heart Circ. Physiol.* **303,** H1319–H1331 (2012).
23. Jimenez-Sanchez, M. *et al.* The Hedgehog signalling pathway regulates autophagy. *Nature Comm.* **3,** 1200 (2012).
24. Wang, Y., Han, C., Lu, L., Magliato, S. & Wu, T. Hedgehog signaling pathway regulates autophagy in human hepatocellular carcinoma cells. *Hepatology* **58,** 995–1010 (2013).
25. Petralia, R. S. *et al.* Sonic hedgehog promotes autophagy in hippocampal neurons. *Biol. Open* **2,** 499–504 (2013).
26. Singh, R. *et al.* Autophagy regulates lipid metabolism. *Nature* **458,** 1131–1135 (2009).
27. Auteri, J. S., Okada, A., Bochaki, V. & Dice, J. F. Regulation of intracellular protein degradation in IMR-90 human diploid fibroblasts. *J. Cell. Physiol.* **115,** 167–174 (1983).
28. Huang, B. Q. *et al.* Isolation and characterization of cholangiocyte primary cilia. *Am. J. Physiol. Gastrointest. Liver Physiol.* **291,** G500–G509 (2006).
29. Marzella, L., Ahlberg, J. & Glaumann, H. Isolation of autophagic vacuoles from rat liver: morphological and biochemical characterization. *J. Cell Biol.* **93,** 144–154 (1982).
30. Wada, K. *et al.* Application of photoacoustic microscopy to analysis of biological components in tissue sections. *Chem. Pharm. Bull. (Tokyo)* **34,** 1688–1693 (1986).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank N. Mizushima and G. Pazour for providing MEFs and I. R. Veland and S. T. Christensen for providing the plasmid for expression of GFP–inversin. We also thank the personnel at the Analytical Imaging Facility for their technical assistance with TEM, J. Kraut and N. Rodriguez for their early assistance in cilia staining optimization and S. Kaushik for critically reviewing this manuscript. This work was supported by grants from the National Institute of Health National Institute on Aging AG031782 and AG038072, National Institute of Diabetes and Digestive and Kidney Diseases DK098408 (to A.M.C.), Institut National de la Santé et de la Recherche Medicale (to P.C. and I.B.), Institut National du Cancer and L'Agence Nationale de la Recherche (to P.C.), and the generous support of Robert and Renee Belfer (to A.M.C.). O.P. was supported by a Basque Government Postdoctoral Fellowship.

Author Contributions O.P. designed and performed most of the experiments, analysed the data and contributed to writing the paper; I.O. and I.B. performed and analysed the experiments related to Hh signalling; B.P. performed the electron microscopy studies and morphometric analysis; S.S. performed the cytosolic vesicle experiments; A.D.-C. assisted with cell culture; P.C. conceived the part of the study related to Hh signalling, provided interpretation to the data and contributed to the writing and revision of the paper; B.H.S. and P.S. set the bases for the rationale of the study, provided feedback in the interpretation of the data and revised the written manuscript; A.M.C. coordinated the study, designed experiments, analysed data and contributed to the writing and revision of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M.C. (ana-maria.cuervo@einstein.yu.edu) or P.S. (peter.satir@einstein.yu.edu).

ARTICLE

doi:10.1038/nature12578

Crystal structure of a Na^+-bound Na^+,K^+-ATPase preceding the E1P state

Ryuta Kanai[1]*, Haruo Ogawa[1]*, Bente Vilsen[2], Flemming Cornelius[2] & Chikashi Toyoshima[1]

Na^+,K^+-ATPase pumps three Na^+ ions out of cells in exchange for two K^+ taken up from the extracellular medium per ATP molecule hydrolysed, thereby establishing Na^+ and K^+ gradients across the membrane in all animal cells. These ion gradients are used in many fundamental processes, notably excitation of nerve cells. Here we describe 2.8 Å-resolution crystal structures of this ATPase from pig kidney with bound Na^+, ADP and aluminium fluoride, a stable phosphate analogue, with and without oligomycin that promotes Na^+ occlusion. These crystal structures represent a transition state preceding the phosphorylated intermediate (E1P) in which three Na^+ ions are occluded. Details of the Na^+-binding sites show how this ATPase functions as a Na^+-specific pump, rejecting K^+ and Ca^{2+}, even though its affinity for Na^+ is low (millimolar dissociation constant). A mechanism for sequential, cooperative Na^+ binding can now be formulated in atomic detail.

Na^+,K^+-ATPase is expressed in all animal cells and is one of the most important members of the P-type ion-translocating ATPases that include Ca^{2+}-ATPase of fast-twitch muscle sarcoplasmic reticulum (SERCA1a) and gastric H^+,K^+-ATPase, among others. Na^+,K^+-ATPase consists of α- and β-subunits and an auxiliary FXYD protein. The α-subunit is the catalytic subunit, consisting of three cytoplasmic domains (A (actuator), N (nucleotide binding) and P (phosphorylation)) and 10 transmembrane helices. Pumping of ions is achieved by alternation between two major conformational states, E1 and E2 (refs 1, 2) (Supplementary Fig. 1 for the reaction cycle). In E1, the cation-binding sites have high affinity for Na^+ and face the cytoplasm; in E2, the cation-binding sites have low affinity for Na^+ but high affinity for K^+ and face the extracellular side (for a review, see, for example, ref. 3). Na^+,K^+-ATPase is essentially a Na^+ pump, as it works even in the absence of K^+. Only Li^+ and H^+ partly substitute for Na^+. Filling of the three transmembrane Na^+-binding sites allows phosphoryl transfer from ATP to an invariant Asp in the α-subunit to occlude three Na^+ in the transmembrane region (denoted as $E1P\cdot3Na^+$). Until now, crystal structures of Na^+,K^+-ATPase have been limited to E2 species[4–7] (see ref. 8 for a recent review) and have provided little information on the Na^+-binding sites. Of particular interest is the third Na^+ site specific to this ATPase[9]. Here we describe a 2.8 Å-resolution crystal structure of Na^+,K^+-ATPase in the E1·aluminium fluoride (AlF_4^-)·ADP·$3Na^+$ form, a stable analogue of the transition state (E1∼P·ADP·$3Na^+$) preceding E1P·$3Na^+$, in which the cytoplasmic gate is at least partly closed (see Supplementary Discussion). Comparison with the crystal structures of an E2·Pi·$2K^+$ analogue of Na^+,K^+-ATPase (E2·MgF_4^{2-}·$2K^+$; Protein Data Bank accession number 2ZXE (ref. 5)) and the corresponding transition state analogue of Ca^{2+}-ATPase[10,11] (E1·AlF_4^-·ADP·$2Ca^{2+}$; Protein Data Bank accession number 2ZBD) shows how this ATPase works as a Na^+-specific pump. The crystal structures also explain how oligomycin, an antibiotic that inhibits Na^+,K^+-ATPase, promotes Na^+ occlusion[12] (reviewed in ref. 13), as detailed in Supplementary Discussion.

Structure of the E1·AlF_4^-·ADP·$3Na^+$ form

Crystals of Na^+,K^+-ATPase in E1·AlF_4^-·ADP·$3Na^+$ from pig kidney were obtained in both the presence and absence of oligomycin and the elucidated protein structures were virtually the same (root mean squared deviation 0.16 Å; Supplementary Fig. 2). We describe the structure derived from the crystals with oligomycin (Fig. 1), as the redundancy of diffraction data is higher (Supplementary Table 1). The asymmetric unit of the crystal consists of two protomers, each comprising α- and β-subunits and FXYD2 protein. Several phospholipid and cholesterol molecules are also identified (Supplementary Fig. 3). The amino (N)-terminal part of the β-subunit, now fully modelled, approaches the P-domain, but not the A-domain as in H^+,K^+-ATPase[14]. The two protomers are very similar (root mean squared deviation 1.10 Å; Supplementary Fig. 4), but the path of the M5 helix and coordination geometries of the three Na^+ in the transmembrane region are slightly different.

The overall structure of the α-subunit, with a closed cytoplasmic headpiece, is very similar to that of SERCA1a in the corresponding state[10,11] (Protein Data Bank accession number 2ZBD; Fig. 2b and Supplementary Figs 1 and 5). It is more upright than E2·MgF_4^{2-}·$2K^+$, as the M5 helix is straighter, and the A-domain is rotated approximately 100° around an axis nearly perpendicular to the membrane (Fig. 2a and Supplementary Figs 6 and 7a). AlF_4^- and ADP, with the adenine ring stacked with Phe 475, cross-link the N- and P-domains through two critical Arg residues[15] (Arg 544 and Arg 685; Fig. 1c). Mg^{2+} is located in the P-domain, coordinated by Asp 369, Asp 710 and AlF_4^-. These features are as in SERCA1a (refs 10, 11). The disposition of the transmembrane helices is also similar, but there are critical differences (Fig. 2b–d). The most conspicuous is the orientation of the extracellular half of the M4 helix (M4E), which is more horizontal in Na^+,K^+-ATPase (Fig. 2c; see Supplementary Discussion for details). The reason for this difference becomes apparent later.

Transmembrane sodium-binding sites

The $|F_{obs}| - |F_{calc}|$ electron density map, before introducing Na^+ into the atomic model, clearly identifies three Na^+ ions in the transmembrane region, nearer to the cytoplasmic surface (Figs 1d and 3a, b), but at slightly different positions in the two protomers (see Supplementary Fig. 8 for composite omit electron density maps; Supplementary Figs 9–11 for details). We place one associated water molecule at site I,

[1]Institute of Molecular and Cellular Biosciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-0032, Japan. [2]Department of Biomedicine, Aarhus University, 8000 Aarhus C, Denmark.
*These authors contributed equally to this work.

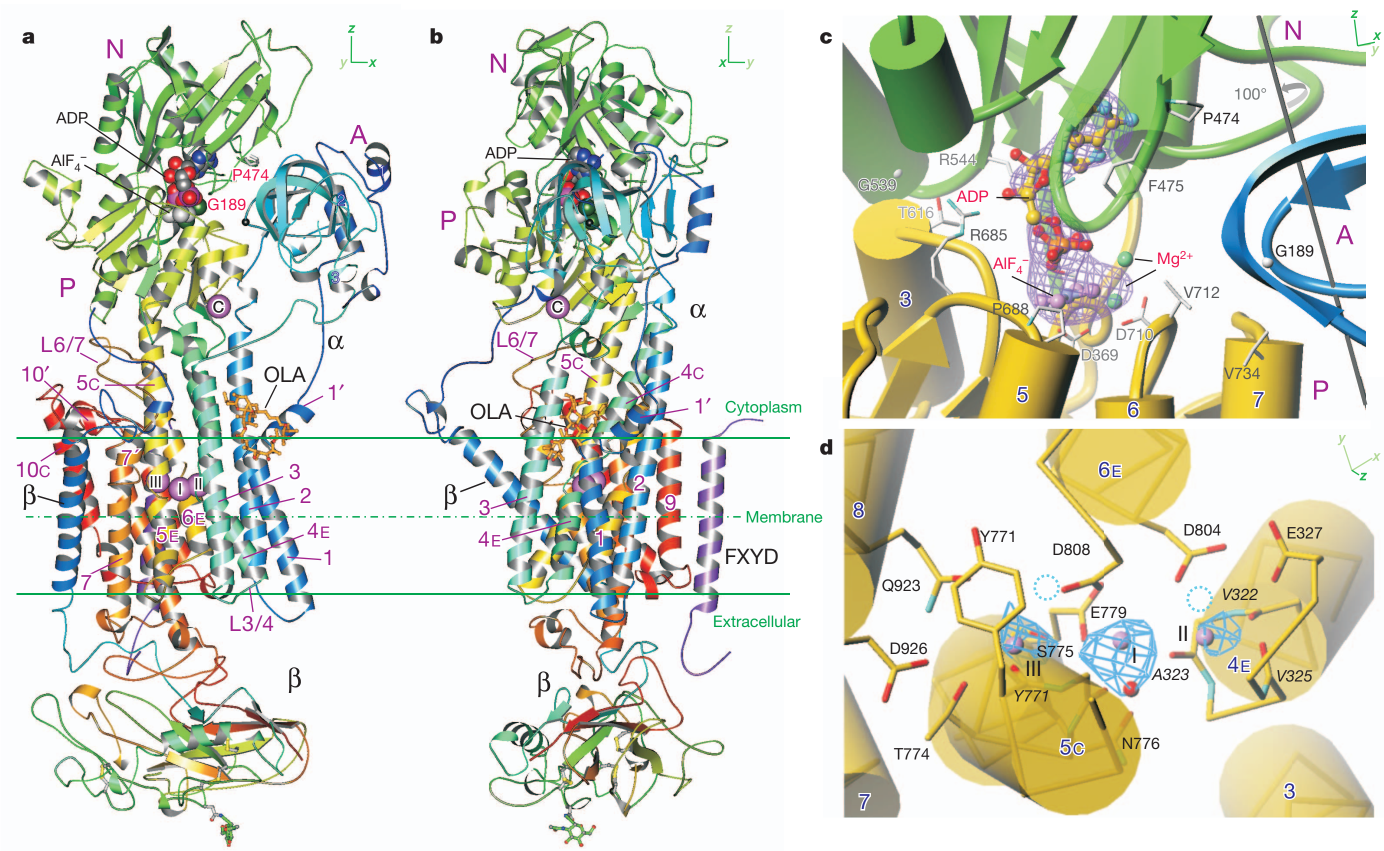


Figure 1 | Crystal structure of Na^+,K^+-ATPase in the transition state analogue E1∼P·ADP·3Na^+. **a**, **b**, Ribbon diagrams viewed in two orthogonal directions. Colour changes gradually between the N terminus (blue) and C terminus (red) for the α- and β-subunits. Purple spheres show bound Na^+ ions (three (I–III) in the transmembrane region and one (C) in the cytoplasmic region[5,41]). Sugars attached to the β-subunit are shown in ball and stick. OLA, oligomycin A. **c**, Details around the nucleotide binding and phosphorylation sites. The axis of rotation of the A-domain is shown. **d**, Transmembrane Na^+-binding sites viewed from the cytoplasmic side. Blue dotted circles indicate the positions of bound Ca^{2+} in SERCA1a in the corresponding state. Residue names in sloping type indicate that the main chain carbonyl contributes to Na^+-coordination. Violet net in **c** represents a σ_A-weighted $|F_{obs}| - |F_{calc}|$ electron density map contoured at 4σ before introducing ADP, AlF_4^- and Mg^{2+} into the atomic model; blue nets in **d** represent the same kind of map at 3σ before introducing three Na^+ (violet spheres) and an associated water molecule (small red sphere) into the atomic model. Transmembrane helices of the α-subunit are numbered (1–10); C or E after the number refers to the cytoplasmic or extracellular half of the helix.

in a slightly different position in each protomer. Partial valence[16] calculated at these locations indicates that these three sites qualify as high-affinity Na^+-binding sites (Fig. 3d), although the coordination geometry is far from an ideal 6 or 7 coordination (Supplementary Table 2), consistent with their millimolar affinity[17]. All three ion-binding sites are largely offset to the M4–M5 side, in marked contrast to the Ca^{2+}-binding sites in SERCA1a[18] (Fig. 1d and Supplementary Fig. 1). Site I Na^+ is located at a position similar to that of the site I K^+ in E2·MgF_4^{2-}·2K^+ (Fig. 3c) but deeper into the interhelix space between M4 and M5; site II Na^+ is located approximately 5 Å closer to the cytoplasmic surface, reflecting the movement of M4E (Fig. 2a). The distance between sites I and II (3.2–3.6 Å) is substantially shorter than that between the two K^+ sites in E2·MgF_4^{2-}·2K^+ (4.1 Å; ref. 5). This distance is compatible with the juxtaposition of two Na^+ (ionic radius = 0.95 Å) but presumably too short for two K^+ (ionic radius = 1.33 Å) and two Ca^{2+} (ionic radius = 0.99 Å). Indeed, 3.6–3.8 Å is the distance between two Na^+ ions forming an ion pair in water[19]. To explore how these binding sites are Na^+ specific, we calculated empirical valence maps[16] (Fig. 3d and Supplementary Fig. 9) assuming cavity radii of 2.4 Å for Na^+ and 2.8 Å for K^+. This cavity size for K^+ correctly identifies K^+ sites in the E2·MgF_4^{2-}·2K^+ crystal structure[5] (Supplementary Fig. 9c).

Sites I and III are juxtaposed at the bottom end of the cytoplasmic half of the M5 helix (M5C), separated by the Ser 775 side chain (Fig. 3b and Supplementary Fig. 10a, b). Here the M5 helix is unwound and kinked owing to Pro 778. Site III is formed at the cytoplasmic end of the unwound part of M5, between the main chain carbonyl of Tyr 771 and Ser 775 amide, as if it were located within the M5 helix (Figs 1d and 2e). Side chains of residues on M6 (Asp 808) and M8 (Gln 923) confine the Na^+ to the unwound space in M5 (Fig. 3a and Supplementary Fig. 11a, b). The aromatic ring of Tyr 771 forms the ceiling of the binding cavity and provides π-electrons for Na^+ binding. A side wall is formed by Thr 774 and Gln 923 side chains and seems rigid because of an extensive hydrogen bonding network involving Tyr 771 (M5), Gln 854 (M7) and Asp 926 (M8). The Ser 775 hydroxyl coordinates both sites I and III Na^+ and is likely to make hydrogen bonds with the Asp 808 carboxyl. The bulky side chain of Tyr 771 fills the interhelix space surrounded by the M5, M6 and M9 helices, and its hydroxyl stabilizes the Thr 807 carbonyl, thereby fixing the unwound part of M6 (Fig. 3a). Involvement of these residues in Na^+ binding is in good agreement with mutagenesis studies[20–25]. Thus, site III seems very restricted in the axial direction (that is, along the M5 helix) and is indeed too small for K^+ (Supplementary Figs 12 and 13).

Of the three Na^+-binding sites, site I, located somewhat outside the M5 helix (Fig. 1d), is the least confined. This is a potential K^+-binding site, as the side chains of Ser 775 and Asn 776 participate in both Na^+ coordination here and K^+ coordination in E2·MgF_4^{2-}·2K^+ (Fig. 3c and Supplementary Fig. 10c). An important difference is that Asp 808 is introduced to the Na^+-binding site by rotation of the unwound part of M6, which is stabilized by Na^+ binding to site III. The presence of Asp 808 forces the binding cation to go deeper into the space between

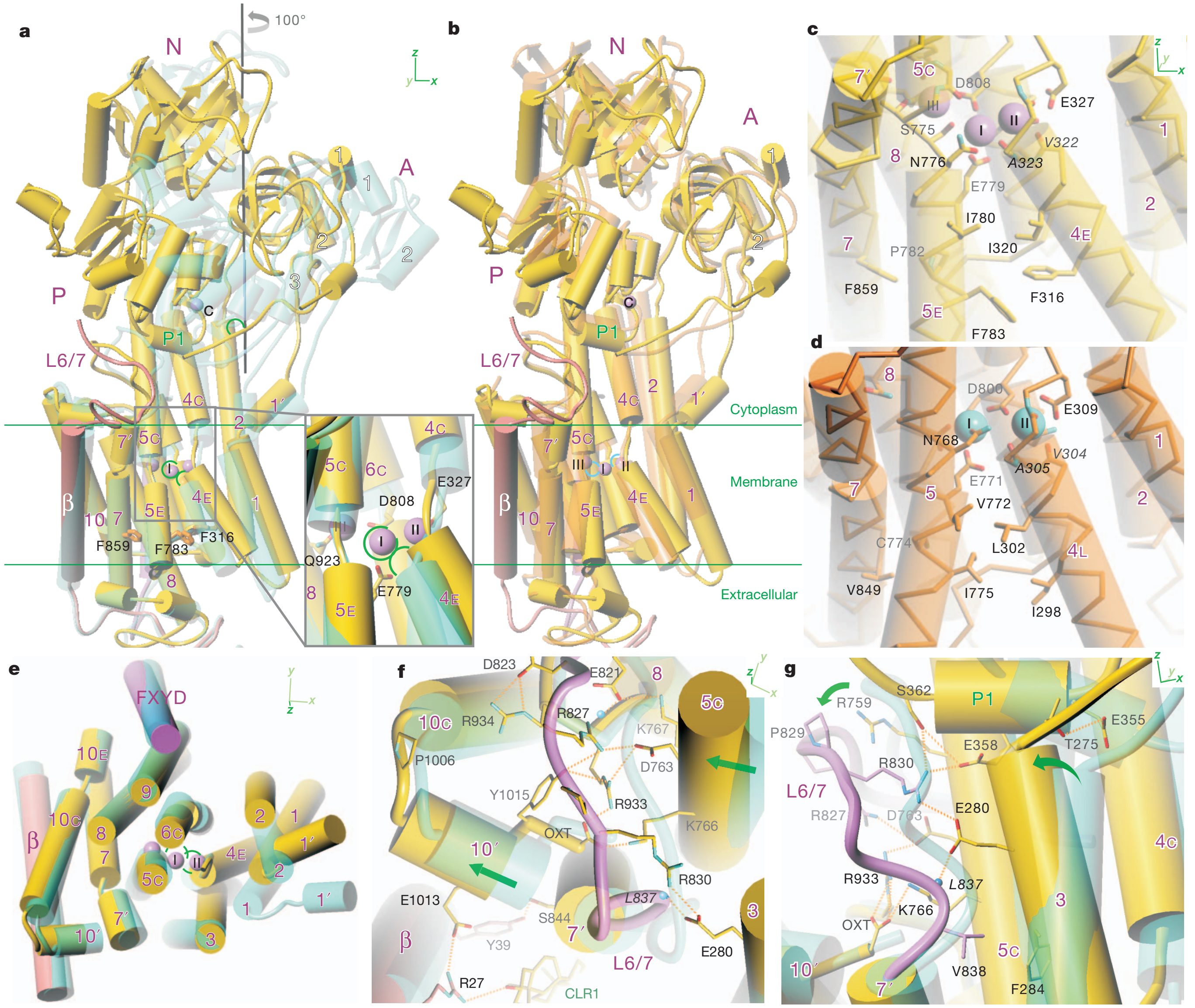


Figure 2 | Comparison of the Na$^+$-bound form of Na$^+$,K$^+$-ATPase with a K$^+$ bound form and with SERCA1a in the corresponding state. Superimposition of Na$^+$,K$^+$-ATPase in E1·AlF$_4^-$·ADP·3Na$^+$ (yellow) with that in E2·MgF$_4^{2-}$·2K$^+$ (cyan; **a**, **e–g**) and with SERCA1a in E1·AlF$_4^-$·ADP·2Ca^{2+} (orange; **b**). **a**, **b**, The whole structure. Inset in **a** shows an enlarged view around the transmembrane cation-binding sites in Na$^+$,K$^+$-ATPase. **c**, **d**, Enlarged views of the transmembrane region of Na$^+$,K$^+$-ATPase (yellow) and SERCA1a (orange) around the cation-binding sites, viewed approximately parallel to the membrane. **e**, Disposition of the transmembrane helices viewed from the cytoplasmic side, approximately perpendicular to the membrane. **f**, The C-terminal region. **g**, The interface between M3 and L6/7. Atomic models are aligned with the M7–M10 helices. The M5 and M7 helices of SERCA1a are contiguous but represented by two cylinders (**b**, **d**). In **a–d**, the M3 helices are removed for clarity. Violet spheres represent Na$^+$, and cyan Ca^{2+}. Green circles represent K$^+$ in E2·MgF$_4^{2-}$·2K$^+$ (**a**, **e**) and blue circles Ca^{2+} in SERCA1a (**b**). In **f** and **g**, the L6/7 loop in E1·AlF$_4^-$·ADP·3Na$^+$ is shown in violet, and arrows indicate the movements in the transition from E2·Pi·2K$^+$ to E1∼P·ADP·3Na$^+$. Dotted lines show likely hydrogen bonds. OXT, terminal carboxyl oxygen.

M4 and M5 towards M3, which, in the case of K$^+$, displaces it from an ideal position (Fig. 3c). Glu 779, which coordinates site II K$^+$ in E2·MgF$_4^{2-}$·2K$^+$, now contributes to site I exclusively. A surprising feature here is that the main chain carbonyl of Ala 323 on M4E, which is dedicated to site II K$^+$ in E2·MgF$_4^{2-}$·2K$^+$, forms a part of site I on the side opposite to Ser 775 and predominantly determines the location of site I (Fig. 3a, b). The coordination by Ala 323 is realized in this state, as M4E is raised by one turn of an α-helix towards the cytoplasm, bringing bulky Ile 320 instead of Ala 323 to interact with Ile 780 (Fig. 2c).

This rearrangement of M4E is achieved primarily through tilting M3 and is closely related to the path of M5C (Fig. 2g and Supplementary Fig. 14). M3, M5C and the L6/7 loop are linked by a complex hydrogen bonding network[5]. As M5C is integrated into the P-domain and the P1 helix running underneath the P-domain is also linked with the cytoplasmic end of M3 (Fig. 2g), M5C affects the path of M3 in several ways. When the M5 helix straightens, the cytoplasmic end of M3 is pulled towards M10 by critical hydrogen bonds with the L6/7 loop. Arg 830 on L6/7 and Glu 280 on M3 are disease-related residues (reviewed in ref. 26); the Arg 830–Gln substitution reduces Na$^+$ affinity fourfold[27]. This pulling of M3 causes a change in inclination (16°) and a large movement at the extracellular end of M3 (Supplementary Fig. 14), as the pivot is formed between Val 838 (L6/7) and Phe 284 (M3) located close to the cytoplasmic end of M3 (Fig. 2g and Supplementary Fig. 14). The rise is transmitted to M4 through a short extracellular loop and makes M4E more horizontal (20°) than in E2·MgF$_4^{2-}$·2K$^+$ (Fig. 2a and Supplementary Fig. 14).

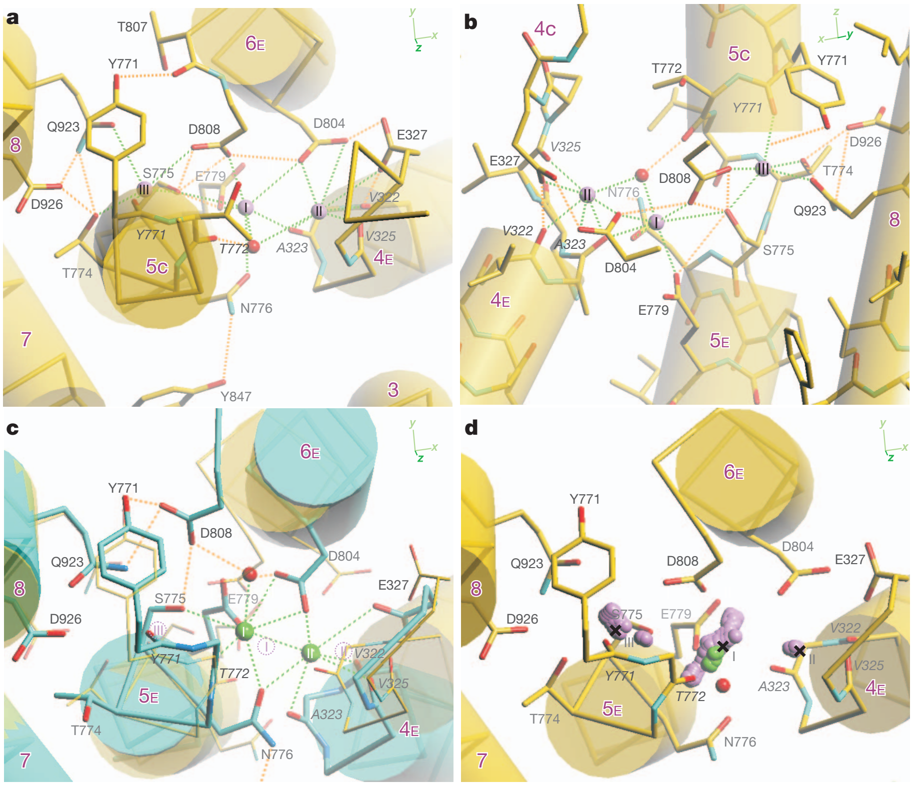


Figure 3 | Transmembrane Na^+-binding sites. **a**, **b**, An atomic model viewed approximately perpendicular to the membrane from the cytoplasm (**a**) and parallel to the membrane from the M6 side (**b**). **c**, Comparison with the K^+-binding sites of Na^+,K^+-ATPase in $E2{\cdot}MgF_4^{2-}{\cdot}2K^+$ (cyan) viewed in the same direction as in **a**; positions of Na^+ are indicated with dotted circles. **d**, A valence map for Na^+ and K^+; grid points with valence ≥ 0.7 are represented with small spheres in violet (Na^+) or green (K^+); crosses indicate the positions of Na^+ in the crystal structure. Residue names in sloping type indicate that the main chain carbonyl contributes to the coordination. Green dotted lines in **a** and **b** indicate likely coordination of Na^+ (≤ 3.7 Å distance), and those in **c** indicate K^+; orange lines, potential hydrogen bonds (provided that carboxyl groups are protonated). Violet spheres represent Na^+, green ones K^+ and red ones water.

This movement of M4E creates site II at a position, with respect to the M5 and M6 helices, distinctly different from that in $E2{\cdot}MgF_4^{2-}{\cdot}2K^+$ (Fig. 3c) and from that in SERCA1a in $E1{\cdot}AlF_4^-{\cdot}ADP{\cdot}2Ca^{2+}$ (Fig. 1d and Supplementary Fig. 1 bottom panel). Nevertheless, the basic coordination geometry is the same: three main chain carbonyls on M4E, a Glu carboxyl (Glu 327 in Na^+,K^+-ATPase) in the unwound part of M4 and one side chain oxygen from an Asp/Asn residue (Asp 804) on M6 coordinate the cation (Na^+). This similarity would indicate that the binding of Na^+ to this site triggers phosphoryl transfer from ATP, although Glu 327 in Na^+,K^+-ATPase and the corresponding Glu 309 in SERCA1a may work differently[28–30]. Here Glu 327 contributes only marginally to the binding (Supplementary Table 2) and is very likely to be protonated (see Supplementary Discussion), in marked contrast to Glu 309 that provides both carboxyl oxygens to Ca^{2+}-coordination[18].

The precise location of site III Na^+ is different in the two protomers (~1.4 Å), consistent with the valence and electron density maps (Supplementary Fig. 9a, b). The position of the valence peak is very sensitive to the path of M5. The deeper position (close to Asp 926) is preferred in protomer B, in which M5 is slightly (~2°) less inclined towards M1 than in protomer A (Supplementary Fig. 15a). As the M5 helix in $E1{\cdot}3Na^+$, immediately after Na^+ binding (Supplementary Fig. 1 for the reaction cycle), would be even less inclined, it is conceivable that site III Na^+ initially binds closer to Asp 926, explaining a large reduction in apparent Na^+ affinity in the Asp 926–Asn mutant[25] (also see Supplementary Discussion). The slight difference in the path of M5 in the two protomers also affects the position of site I Na^+ and the associated water molecule. The valence map shows that K^+ binding is possible at a position above site I Na^+ (Fig. 3d and Supplementary Fig. 9a, b), but presumably not in $E1{\cdot}3Na^+$, as the coordinating residues are likely to be shifted towards M8.

These observations highlight the importance of movements of M5C and their control. The valence maps show that sites II and III are too narrow for K^+. However, if the M5 helix straightens like that in the $E1{\cdot}2Ca^{2+}$ crystal of SERCA1a (Protein Data Bank accession number 1SU4)[18], K^+ binding to site III seems possible (Supplementary Fig. 15b). This requires only about 10° more inclination. To establish selectivity for Na^+, the orientation of M5C has to be critically controlled, much more stringently than in SERCA1a. Interestingly, Na^+,K^+-ATPase has a unique carboxy (C)-terminal extension, the deletion of which decreases Na^+ affinity[4,31–34]. This part consists of a short helix (M10′) connected to the M10 helix with a short loop. Superimposition of the segment on that of the $E2{\cdot}MgF_4^{2-}{\cdot}2K^+$ structure (Fig. 2f) shows that M10′ moves approximately 2.5 Å while the M10 helix remains nearly fixed. Thus, the C-terminal extension could work as a brake or damper that prevents excessive movement of the M5 helix, together with the L6/7 loop, M7′ and M8 linked by an extensive hydrogen bonding network; similarly, the movement of M7′ is controlled in part by the β-subunit (Fig. 2f and Supplementary Fig. 16). Deletion of the C-terminal residues will allow larger movements of M5, which will decrease Na^+ affinity and allow binding of K^+ to site III. As K^+ binding to site III would hinder closure of the cytoplasmic headpiece, ATPase activity would decrease at high K^+ concentrations. Such behaviour is in fact observed on the deletion of the C-terminal five residues[4,31] and on substitutions of Arg 933 on M8 (ref. 31), a key residue in the hydrogen bonding network (Fig. 2f).

Sequence of binding

We can now provide a possible scenario for sequential[35,36], cooperative[37] binding of the three Na^+ ions. A continuous cavity large enough for the passage of Na^+ (but not for K^+) from the cytoplasm to site III

can be formed if Ser 775 changes its side chain conformation (Fig. 4a and Supplementary Fig. 13a). We do not see other possible pathways to site III unless very large conformational changes are assumed. Thus, the crystal structure predicts that the first Na^+ binds to site III and the last to site II. Yet, for the first Na^+ to be able to reach site III, sites I and II need to be unformed, and Na^+ binding to site III must create site I.

First, the headpiece must open (aided by ATP) and K^+ must be released. In the E1 state thus realized, K^+ can still bind and reverse the ATPase into E2·2K^+. In E2·MgF_4^{2-}·2K^+, Na^+-coordinating residues at site III (that is, Tyr 771, Ser 775, Asp 808 and Gln 923) are all tied by hydrogen bonds and two bound K^+ (Fig. 3c and Supplementary Figs 10c and 11c), leaving no space for Na^+. Release of K^+ will set the unwound part of M6 free to rotate and Ser 775 to change its side-chain orientation, thereby allowing the first Na^+ to reach site III. Na^+ binding to site III fixes the Ser 775 side chain and the unwound part of M6, making the Asp 804 and Asp 808 carboxyls available for site I Na^+ (Fig. 4b). This will be the first cooperative step, bringing in three more oxygen atoms to site I. For the second Na^+ to reach site I, site II must remain unformed. In this state, M4E is likely to take a position similar to the corresponding segment (the lumenal half of M4, M4L) of SERCA1a in E1·Mg^{2+} (refs 38, 39), a state ready to accept incoming Ca^{2+} to site I with site II completely disrupted owing to displaced M4L. Ca^{2+}-binding to site I will bring M4L towards M6 so that site II becomes properly formed[38]. In Na^+,K^+-ATPase, the position of M4E is, initially, largely determined by the binding of Na^+ to site III as described (Supplementary Fig. 14). The binding of Na^+ to site I would bring the Ala 323 carbonyl on the cytoplasmic end of M4E to where it coordinates both site I and site II Na^+ (Fig. 4c). This repositioning of M4E will form site II properly. Thus, binding of Na^+ to site III confers Na^+ specificity to sites I and II, and completely different mechanisms are used to link sites III and I, and sites I and II. The first Na^+ binding stabilizes an unwound portion of a helix to bring carboxyls into the second binding site; the second Na^+ binding repositions one end of a helix to form the third binding site properly. The binding of the third Na^+ (Fig. 4d) will change the path of M4C so that the P-domain can bend and accept Mg^{2+} and phosphate transferred from ATP[10,38]. In this way the mechanism of the cooperative Na^+ binding and activation of phosphoryl transfer can be explained.

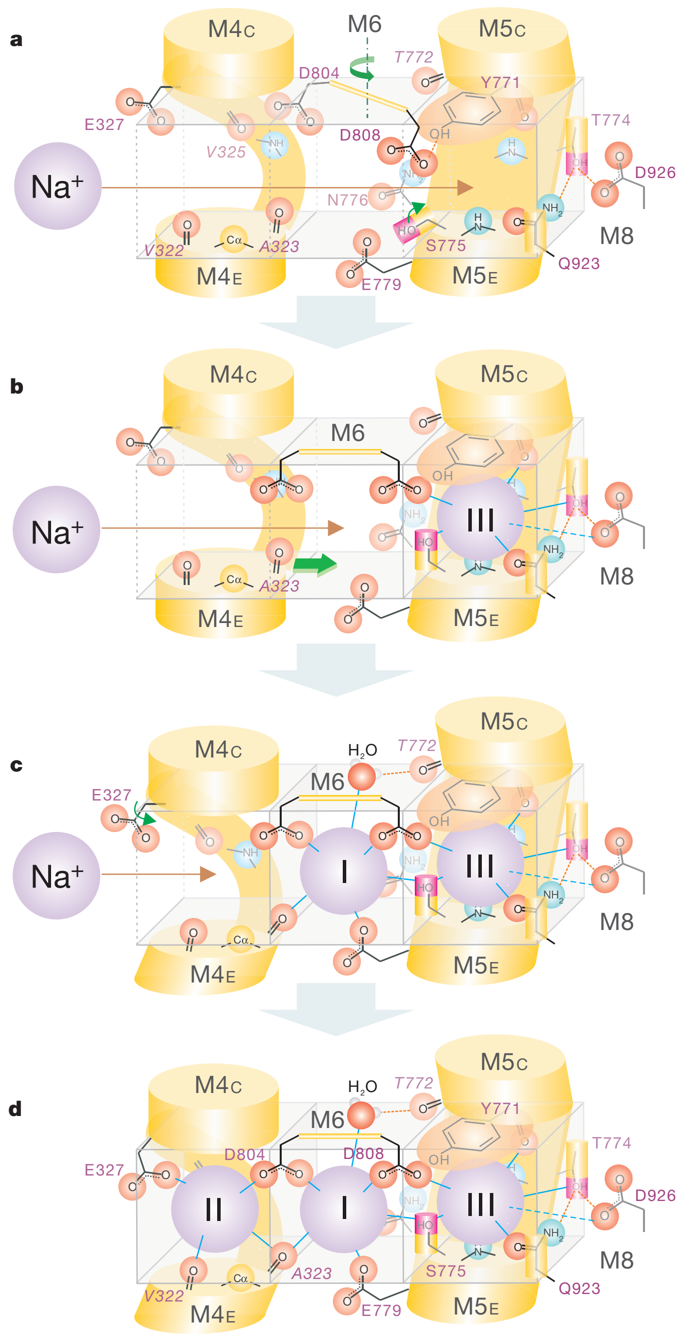


Figure 4 | A proposed model for binding sequence of three Na^+. A cartoon showing the transmembrane binding sites viewed in a way similar to Fig. 3b. Thin blue lines represent coordination of Na^+.

Conclusion

It has been an enigma how low-affinity (millimolar dissociation constant in the presence of Mg^{2+}; ref 17) transmembrane Na^+-binding sites can effectively reject K^+ and Ca^{2+} and make Na^+,K^+-ATPase work as a Na^+-pump, at the same time countertransporting K^+ using some of the same binding residues. The crystal structure presented here suggests that a threefold strategy is used. The Na^+-binding sites are highly constrained to fit only small Na^+ ions. In all three sites, cavity walls are rigid with an extensive hydrogen bonding network, and main chain carbonyl groups located on ends of helices are used for coordination. This must be at least a part of the reason why the three Na^+-binding sites are largely offset towards M4 and M5 (Fig. 1d). Second, the ion-binding sites are juxtaposed so tightly that only small monovalent ions can bind simultaneously (applicable to sites I and II). This means that the stoichiometry of three Na^+ ions transported per ATP molecule hydrolysed is not only for efficiency but also for specificity. Third, the first binding site (that is, site III) is placed on a particularly restrictive and critical hinge so that only the right-sized ion can promote productive reactions upon binding. For phosphoryl transfer and occlusion, proper positioning of the P-domain that coordinates two other cytoplasmic domains is critical[10], and therefore that of the M5C helix. Na^+ promotes this but not K^+ even if it happened to bind. This is reminiscent of allosteric control by a ligand that binds to the hinge between two domains and thereby controls catalysis. The mechanism is thus significantly different from that used in SERCA1a[18], where side chains are used exclusively in binding the first Ca^{2+}.

METHODS SUMMARY

Na^+,K^+-ATPase was isolated from pig kidney outer medulla and purified by mild SDS treatment followed by isopycnic zonal centrifugation[40]. Crystals were obtained by vapour diffusion at 10 °C. Crystal structures were determined by molecular replacement starting from a homology model based on the corresponding structure of Ca^{2+}-ATPase (Protein Data Bank accession number: 2ZBD). The atomic model, consisting of residues 20–1016 of the α-subunit, the complete β-subunit and 15–49 of the FXYD2 protein, five phosphatidylcholine and three cholesterol molecules per protomer, was refined to R_{work} of 0.268 and R_{free} of 0.297 (oligomycin-bound) and to R_{work} of 0.263 and R_{free} of 0.299 (oligomycin-free) at 2.8-Å resolution. Diffraction data and refinement statistics are summarized in Supplementary Table 1. Partial valence was calculated according to ref. 16.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 May; accepted 16 August 2013.
Published online 2 October 2013.

1. Albers, R. W. Biochemical aspects of active transport. *Annu. Rev. Biochem.* **36,** 727–756 (1967).
2. Post, R. L., Hegyvary, C. & Kume, S. Activation by adenosine triphosphate in the phosphorylation kinetics of sodium and potassium ion transport adenosine triphosphatase. *J. Biol. Chem.* **247,** 6530–6540 (1972).
3. Apell, H. J. & Karlish, S. J. Functional properties of Na,K-ATPase, and their structural implications, as detected with biophysical techniques. *J. Membr. Biol.* **180,** 1–9 (2001).
4. Morth, J. P. *et al.* Crystal structure of the sodium–potassium pump. *Nature* **450,** 1043–1049 (2007).
5. Shinoda, T., Ogawa, H., Cornelius, F. & Toyoshima, C. Crystal structure of the sodium–potassium pump at 2.4 Å resolution. *Nature* **459,** 446–450 (2009).
6. Ogawa, H., Shinoda, T., Cornelius, F. & Toyoshima, C. Crystal structure of the sodium-potassium pump (Na^+,K^+-ATPase) with bound potassium and ouabain. *Proc. Natl Acad. Sci. USA* **106,** 13742–13747 (2009).
7. Laursen, M., Yatime, L., Nissen, P. & Fedosova, N. U. Crystal structure of the high-affinity Na^+,K^+-ATPase–ouabain complex with Mg^{2+} bound in the cation binding site. *Proc. Natl Acad. Sci. USA* **110,** 10958–10963 (2013).
8. Toyoshima, C., Kanai, R. & Cornelius, F. First crystal structures of Na^+,K^+-ATPase: new light on the oldest ion pump. *Structure* **19,** 1732–1738 (2011).
9. Morth, J. P. *et al.* A structural overview of the plasma membrane Na^+,K^+-ATPase and H^+-ATPase ion pumps. *Nature Rev. Mol. Cell Biol.* **12,** 60–70 (2011).
10. Toyoshima, C. & Mizutani, T. Crystal structure of the calcium pump with a bound ATP analogue. *Nature* **430,** 529–535 (2004).
11. Sørensen, T. L., Møller, J. V. & Nissen, P. Phosphoryl transfer and calcium ion occlusion in the calcium pump. *Science* **304,** 1672–1675 (2004).
12. Esmann, M. & Skou, J. C. Occlusion of Na^+ by the Na,K-ATPase in the presence of oligomycin. *Biochem. Biophys. Res. Commun.* **127,** 857–863 (1985).
13. Glynn, I. M. in *The Enzymes of Biological Membranes* Vol. 3 (ed. A. N. Martonosi) 35–114 (1985). New York and London: Plenum.
14. Abe, K., Tani, K., Nishizawa, T. & Fujiyoshi, Y. Inter-subunit interaction of gastric H^+,K^+-ATPase prevents reverse reaction of the transport cycle. *EMBO J.* **28,** 1637–1643 (2009).
15. Jacobsen, M. D., Pedersen, P. A. & Jørgensen, P. L. Importance of Na,K-ATPase residue α1-Arg^{544} in the segment Arg^{544}–Asp^{567} for high-affinity binding of ATP, ADP, or MgATP. *Biochemistry* **41,** 1451–1456 (2002).
16. Nayal, M. & Di Cera, E. Valence screening of water in protein crystals reveals potential Na^+ binding sites. *J. Mol. Biol.* **256,** 228–234 (1996).
17. Pintschovius, J., Fendler, K. & Bamberg, E. Charge translocation by the Na^+/K^+-ATPase investigated on solid supported membranes: cytoplasmic cation binding and release. *Biophys. J.* **76,** 827–836 (1999).
18. Toyoshima, C., Nakasako, M., Nomura, H. & Ogawa, H. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* **405,** 647–655 (2000).
19. Guàrdia, E., Rey, R. & Padró, J. A. Na^+–Na^+ and Cl^-–Cl^- ion pairs in water: mean force potentials by constrained molecular dynamics. *J. Chem. Phys.* **95,** 2823–2831 (1991).
20. Vilsen, B. Mutant Glu781→Ala of the rat kidney Na^+,K^+-ATPase displays low cation affinity and catalyzes ATP hydrolysis at a high rate in the absence of potassium ions. *Biochemistry* **34,** 1455–1463 (1995).
21. Vilsen, B., Ramlov, D. & Andersen, J. P. Functional consequences of mutations in the transmembrane core region for cation translocation and energy transduction in the Na^+,K^+-ATPase and the SR Ca^{2+}-ATPase. *Ann. NY Acad. Sci.* **834,** 297–309 (1997).
22. Pedersen, P. A., Nielsen, J. M., Rasmussen, J. H. & Jørgensen, P. L. Contribution to Tl^+, K^+, and Na^+ binding of Asn776, Ser775, Thr774, Thr772, and Tyr771 in cytoplasmic part of fifth transmembrane segment in α-subunit of renal Na,K-ATPase. *Biochemistry* **37,** 17818–17827 (1998).
23. Kuntzweiler, T. A., Arguello, J. M. & Lingrel, J. B. Asp804 and Asp808 in the transmembrane domain of the Na,K-ATPase α subunit are cation coordinating residues. *J. Biol. Chem.* **271,** 29682–29687 (1996).
24. Imagawa, T., Yamamoto, T., Kaya, S., Sakaguchi, K. & Taniguchi, K. Thr-774 (transmembrane segment M5), Val-920 (M8), and Glu-954 (M9) are involved in Na^+ transport, and Gln-923 (M8) is essential for Na,K-ATPase activity. *J. Biol. Chem.* **280,** 18736–18744 (2005).
25. Einholm, A. P., Toustrup-Jensen, M. S., Holm, R., Andersen, J. P. & Vilsen, B. The rapid-onset dystonia parkinsonism mutation D923N of the Na^+, K^+-ATPase α3 isoform disrupts Na^+ interaction at the third Na^+ site. *J. Biol. Chem.* **285,** 26245–26254 (2010).
26. Morth, J. P. *et al.* The structure of the Na^+,K^+-ATPase and mapping of isoform differences and disease-related mutations. *Phil. Trans. R. Soc. B* **364,** 217–227 (2009).
27. Schack, V. R., Holm, R. & Vilsen, B. Inhibition of phosphorylation of Na^+,K^+-ATPase by mutations causing familial hemiplegic migraine. *J. Biol. Chem.* **287,** 2191–2202 (2012).
28. Jewell-Motz, E. A. & Lingrel, J. B. Site-directed mutagenesis of the Na,K-ATPase: consequences of substitutions of negatively-charged amino acids localized in the transmembrane domains. *Biochemistry* **32,** 13523–13530 (1993).
29. Vilsen, B. Glutamate 329 located in the fourth transmembrane segment of the α-subunit of the rat kidney Na^+,K^+-ATPase is not an essential residue for active transport of sodium and potassium ions. *Biochemistry* **32,** 13340–13349 (1993).
30. Clarke, D. M., Loo, T. W., Inesi, G. & MacLennan, D. H. Location of high affinity Ca^{2+}-binding sites within the predicted transmembrane domain of the sarcoplasmic reticulum Ca^{2+}-ATPase. *Nature* **339,** 476–478 (1989).
31. Toustrup-Jensen, M. S. *et al.* The C terminus of Na^+,K^+-ATPase controls Na^+ affinity on both sides of the membrane through Arg935. *J. Biol. Chem.* **284,** 18715–18725 (2009).
32. Yaragatupalli, S., Olivera, J. F., Gatto, C. & Artigas, P. Altered Na^+ transport after an intracellular α-subunit deletion reveals strict external sequential release of Na^+ from the Na/K pump. *Proc. Natl Acad. Sci. USA* **106,** 15507–15512 (2009).
33. Meier, S., Tavraz, N. N., Durr, K. L. & Friedrich, T. Hyperpolarization-activated inward leakage currents caused by deletion or mutation of carboxy-terminal tyrosines of the Na^+/K^+-ATPase α subunit. *J. Gen. Physiol.* **135,** 115–134 (2010).
34. Vedovato, N. & Gadsby, D. C. The two C-terminal tyrosines stabilize occluded Na/K pump conformations containing Na or K ions. *J. Gen. Physiol.* **136,** 63–82 (2010).
35. Schulz, S. & Apell, H. J. Investigation of ion binding to the cytoplasmic binding sites of the Na,K-pump. *Eur. Biophys. J.* **23,** 413–421 (1995).
36. Cornelius, F. & Skou, J. C. The sided action of Na^+ on reconstituted shark Na^+/K^+-ATPase engaged in Na^+–Na^+ exchange accompanied by ATP hydrolysis. II. Transmembrane allosteric effects on Na^+ affinity. *Biochim. Biophys. Acta* **944,** 223–232 (1988).
37. Schneeberger, A. & Apell, H. J. Ion selectivity of the cytoplasmic binding sites of the Na,K-ATPase: II. Competition of various cations. *J. Membr. Biol.* **179,** 263–273 (2001).
38. Toyoshima, C. *et al.* Crystal structures of the calcium pump and sarcolipin in the Mg^{2+}-bound E1 state. *Nature* **495,** 260–264 (2013).
39. Winther, A. M. *et al.* The sarcolipin-bound calcium pump stabilizes calcium sites exposed to the cytoplasm. *Nature* **495,** 265–269 (2013).
40. Jørgensen, P. L. Purification and characterization of (Na^+ + K^+)-ATPase. 3. Purification from the outer medulla of mammalian kidney after selective removal of membrane components by sodium dodecylsulphate. *Biochim. Biophys. Acta* **356,** 36–52 (1974).
41. Schack, V. R. *et al.* Identification and function of a cytoplasmic K^+ site of the Na^+, K^+ -ATPase. *J. Biol. Chem.* **283,** 27982–27990 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Hasegawa and H. Okumura of the Japan Synchrotron Radiation Research Institute (JASRI) for data collection at BL41XU of SPring-8, and J. Petersen, A. M. Beck Sørensen and A. Lillevang for technical assistance with membrane purification. We are grateful to D. B. McIntosh for help in improving the manuscript, and J. Tsueda for help in preparing figures. This work is a part of a long-term project (2009B0025) at SPring-8 and was supported by a Specially Promoted Project Grant from the Ministry of Education, Culture, Sports, Science and Technology of Japan (to C.T.) as well as by research grants from the Lundbeck Foundation (to B.V.), the Novo Nordisk Foundation, Fabrikant Vilhelm Pedersen og Hustrus Legat (to B.V. and F.C.), the Danish Medical Research Council (to B.V. and F.C.) and the FI/Danish-Japanese cooperation program from the Danish Agency for Science Technology and Innovation (to F.C.).

Author Contributions C.T., B.V. and F.C. planned and supervised the study. B.V. and F.C. prepared purified membranes containing Na^+,K^+-ATPase. R.K. and C.T. crystallized the ATPase; R.K., H.O. and C.T. collected diffraction data, determined the structure and prepared figures. C.T., B.V. and F.C. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Atomic coordinates and structure factors for the reported crystal structures are deposited in the Protein Data Bank under accession numbers 3WGV (with oligomycin) and 3WGU (without oligomycin). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.T. (ct@iam.u-tokyo.ac.jp).

LETTER

doi:10.1038/nature12522

Evidence for a new nuclear 'magic number' from the level structure of ^{54}Ca

D. Steppenbeck[1], S. Takeuchi[2], N. Aoi[3], P. Doornenbal[2], M. Matsushita[1], H. Wang[2], H. Baba[2], N. Fukuda[2], S. Go[1], M. Honma[4], J. Lee[2], K. Matsui[5], S. Michimasa[1], T. Motobayashi[2], D. Nishimura[6], T. Otsuka[1,5], H. Sakurai[2,5], Y. Shiga[7], P.-A. Söderström[2], T. Sumikama[8], H. Suzuki[2], R. Taniuchi[5], Y. Utsuno[9], J. J. Valiente-Dobón[10] & K. Yoneda[2]

Atomic nuclei are finite quantum systems composed of two distinct types of fermion—protons and neutrons. In a manner similar to that of electrons orbiting in an atom, protons and neutrons in a nucleus form shell structures. In the case of stable, naturally occurring nuclei, large energy gaps exist between shells that fill completely when the proton or neutron number is equal to 2, 8, 20, 28, 50, 82 or 126 (ref. 1). Away from stability, however, these so-called 'magic numbers' are known to evolve in systems with a large imbalance of protons and neutrons. Although some of the standard shell closures can disappear, new ones are known to appear[2,3]. Studies aiming to identify and understand such behaviour are of major importance in the field of experimental and theoretical nuclear physics. Here we report a spectroscopic study of the neutron-rich nucleus ^{54}Ca (a bound system composed of 20 protons and 34 neutrons) using proton knockout reactions involving fast radioactive projectiles. The results highlight the doubly magic nature of ^{54}Ca and provide direct experimental evidence for the onset of a sizable subshell closure at neutron number 34 in isotopes far from stability.

The shell structure of the atomic nucleus was first successfully described more than 60 years ago[1]. However, the question of how robust the standard magic numbers are in unstable nuclei with a large excess of neutrons—often referred to as 'exotic' nuclei—has been one of the main driving forces behind recent nuclear structure studies that focus on changes in the shell structure, called 'shell evolution'. A noteworthy example is the disappearance of the $N = 28$ (neutron number 28) standard magic number in ^{42}Si (ref. 4), a nucleus that lies far from the stable isotopes on the Segrè chart. On the contrary, exotic oxygen isotopes[3] provide evidence for the onset of a new shell closure at $N = 16$, one that is not observed in stable nuclei. In both cases, the tensor force, a non-central component of the nuclear force, has a key role in describing the experimental spectra[5].

The region of the Segrè chart around exotic calcium isotopes has also contributed valuable input to the understanding of nuclear shell evolution over recent years owing to experimental advances. Enhanced excitation energies of first $J^{\Pi} = 2^+$ states (spin, J; parity, Π) and reduced γ-ray transition probabilities, which are good indicators of nuclear shell gaps, for ^{52}Ca (refs 6, 7), ^{54}Ti (refs 8, 9) and ^{56}Cr (refs 10, 11) provide substantial evidence for the onset of a sizable energy gap at $N = 32$. This result was recently confirmed by high-precision mass measurements on neutron-rich Ca isotopes[12]. In the framework of tensor-force-driven shell evolution[5], the $N = 32$ subshell closure is a direct consequence of the weakening of the attractive nucleon–nucleon interaction between protons (π) and neutrons (ν) in the $\pi f_{7/2}$ and $\nu f_{5/2}$ single-particle orbitals (SPOs) as the number of protons in the $\pi f_{7/2}$ SPO is reduced and the magnitude of the $\pi f_{7/2}$–$\nu f_{5/2}$ energy gap increases (Fig. 1a–c).

A question that has been asked frequently over recent years is whether or not the onset of another subshell gap occurs in exotic $N = 34$ isotones, which was suggested qualitatively more than a decade ago[13] on the basis of the general properties of nuclear forces. The onset of an appreciable subshell closure at $N = 34$ is illustrated in Fig. 1d, indicating an energy gap between the $\nu p_{1/2}$ and $\nu f_{5/2}$ SPOs in ^{54}Ca that is comparable to the separation of the $\nu p_{3/2}$ and $\nu p_{1/2}$ spin–orbit partners, which is also implied by recent theoretical results; see, for example, ref. 14. We stress, however, that no $N = 34$ subshell closure was reported in the experimental investigations of ^{56}Ti (refs 9, 15) or ^{58}Cr (refs 11, 16), and notable doubt on this magic number for Ca isotopes has been raised[17,18]. Indeed, as indicated in Fig. 2a, theoretical predictions of the energy of the first $J^{\Pi} = 2^+$ state for ^{54}Ca vary considerably, ranging from $\sim$1 MeV in some cases to as high as $\sim$4 MeV in others[14–16,19–24], despite exhibiting close agreement for lighter isotopes; for example, the predictions of the same theories lie within only 0.4 MeV of the empirical result for ^{52}Ca. Such stark discrepancies at $N = 34$ reflect the need for direct experimental input on the matter.

To address this issue, we report on an experimental study of ^{54}Ca to clarify the strength of the $N = 34$ subshell gap in nuclei farther from stability. The energies of nuclear excited states were investigated using proton knockout reactions involving ^{55}Sc and ^{56}Ti projectiles on a Be target at the Radioactive Isotope Beam Factory, Japan, operated by the RIKEN Nishina Center and the Center for Nuclear Study, University of Tokyo. Experimental details are provided in Methods Summary. Particle identification plots indicating the radioactive species transported through the BigRIPS separator and ZeroDegree spectrometer[25], which were used to select and tag radioactive beam projectiles and reaction products, are presented in Fig. 3a and Fig. 3b, respectively. We emphasize that the intensity of the radioactive beam reported here, which was critical to the success of the experiment, is unique to the Radioactive Isotope Beam Factory. Excited-state energies were deduced using the technique of in-beam γ-ray spectroscopy.

The γ-rays measured in coincidence with ^{54}Ca projectiles produced through the one- and two-proton knockout reaction channels are presented in Fig. 4a. The γ-ray energies measured in the laboratory frame of reference have been corrected for Doppler shifts, and so the transitions appear at the energies they would in the rest frame of the nucleus. The most intense γ-ray line in the ^{54}Ca spectrum, the peak at 2,043(19) keV (error, 1 s.d.) in Fig. 4a, is assigned as the transition from the first 2^+ state (2_1^+) to the 0^+ ground state. In addition, two weaker transitions are located at 1,656(20) and, respectively, 1,184(24) keV. Figure 4b shows a γ-ray spectrum obtained with the condition of a prompt coincidence ($\lesssim$10 ns) with the 2,043-keV γ-ray, indicating that the weaker transitions were emitted in decay sequences involving the $2_1^+ \rightarrow 0^+$ ground-state transition. On the basis of the γ-ray relative intensities, the 1,656-keV transition is proposed to depopulate a level at 3,699(28) keV, as presented in the ^{54}Ca level scheme in the lower-right section of Fig. 4a. Placement of the 1,184-keV transition in the

[1]Center for Nuclear Study, University of Tokyo, Hongo, Bunkyo, Tokyo 113-0033, Japan. [2]RIKEN Nishina Center, 2-1, Hirosawa, Wako, Saitama 351-0198, Japan. [3]Research Center for Nuclear Physics, University of Osaka, Ibaraki, Osaka 567-0047, Japan. [4]Center for Mathematical Sciences, Aizu University, Aizu-Wakamatsu, Fukushima 965-8580, Japan. [5]Department of Physics, University of Tokyo, Hongo, Bunkyo, Tokyo 113-0033, Japan. [6]Department of Physics, Tokyo University of Science, Noda, Chiba 278-8510, Japan. [7]Department of Physics, Rikkyo University, Toshima, Tokyo 171-8501, Japan. [8]Department of Physics, Tohoku University, Sendai, Miyagi 980-8578, Japan. [9]Japan Atomic Energy Agency, Tokai, Ibaraki 319-1195, Japan. [10]Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Legnaro, Legnaro 35020, Italy.

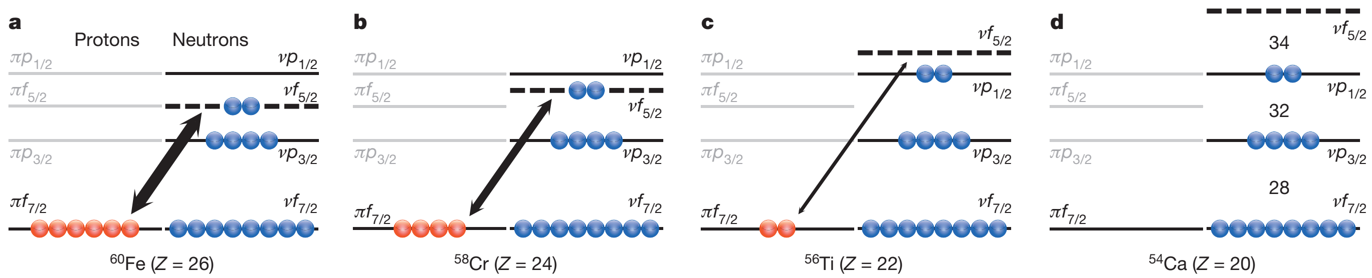


Figure 1 | Schematic illustration highlighting the attractive interaction between the proton $\pi f_{7/2}$ and neutron $\nu f_{5/2}$ single-particle orbitals for $N = 34$ isotones. a–c, As protons are removed from the $\pi f_{7/2}$ orbital (from ^{60}Fe (**a**) through ^{58}Cr (**b**) to ^{56}Ti (**c**)), the strength of the π–ν interaction decreases, as represented by the decreasing width of the diagonal arrows, causing the $\nu f_{5/2}$ orbital to shift up in energy relative to the $\nu p_{3/2}$–$\nu p_{1/2}$ spin–orbit partners. Consequently, a sizable subshell closure presents itself at $N = 32$ in isotopes far from stability. **d**, An additional subshell closure at $N = 34$ for ^{54}Ca is possible. The $\nu f_{5/2}$ SPO is indicated as a bold dashed line to guide the eye.

level scheme is uncertain owing to ambiguity in coincidence measurements between the 1,656- and 1,184-keV lines. However, the appearance of a relatively low-energy peak ($\lesssim$0.5 MeV) in the energy-gated spectrum in Fig. 4b provides a further hint of the level structure. We note that this peak is obscured in the spectrum of Fig. 4a because Bremsstrahlung radiation is present. Although the precise energy of the peak is not reported here, owing to ambiguity caused by detector thresholds, the peak appears in the spectrum at an energy corresponding to the energy difference between the 1,656- and 1,184-keV transitions, which is 472(31) keV. Thus, it is probable that the 1,184- and ~472-keV transitions form a cascade that starts from the 3,699-keV level and runs in parallel to the 1,656-keV γ-ray, although the ordering of the 1,184- and ~472-keV transitions in the decay sequence cannot be specified here.

The excitation energies of 2_1^+ states, $E(2_1^+)$, are presented in Fig. 2b for Ca isotopes. Typically, peaks in $E(2_1^+)$ systematics along isotopic chains indicate the presence of large nuclear shell gaps, although the correlation energy can also influence $E(2_1^+)$ in some instances. Indeed, in the case of ^{48}Ca, the large $E(2_1^+)$ value reflects the strength of the standard magic numbers $Z = 20$ (proton number 20) and $N = 28$ in stable nuclei. For ^{52}Ca, the onset of the $N = 32$ subshell closure[6,7,12] is highlighted by the large increase in $E(2_1^+)$ relative to its even–even (even-N and even-Z) neighbour ^{50}Ca. The result of the present work indicates that $E(2_1^+)$ for ^{54}Ca is comparable to that for ^{52}Ca, thus providing direct experimental evidence for the doubly magic nature of ^{54}Ca.

Moreover, in Fig. 2c, $E(2_1^+)$ systematics for the $N = 30$, 32 and 34 isotonic chains are presented for even–even nuclei from Ar ($Z = 18$) to Ge ($Z = 32$). A comparison of the energies for Ca and Ni along each isotonic chain is emphasized, because these two species are host to magic proton cores with $Z = 20$ and 28, respectively. For the $N = 30$ isotones, we note that $E(2_1^+)$ for ^{58}Ni lies 0.4 MeV above its Ca counterpart. In the case of ^{60}Ni and ^{52}Ca, however, which fall on the $N = 32$ line in Fig. 2c, $E(2_1^+)$ becomes strongly enhanced for Ca relative to Ni, indicating the additional strength of the $N = 32$ subshell gap at $Z = 20$. Similarly, a large increase in $E(2_1^+)$ for ^{54}Ca relative to other $N = 34$ isotones provides further evidence for the new subshell closure in Ca isotopes at this neutron number.

Also plotted in Fig. 2b are the energies of first $J^{\Pi} = 3^-$ states, $E(3_1^-)$, for even–even Ca isotopes alongside the 3,699-keV level reported in the present work. Discussions of the systematics of $E(3_1^-)$ along the Ca isotopic chain and specific details of the nature of those excitations, which are understood as nucleon cross-shell excitations, are provided in refs 7, 26. Because the data point for ^{54}Ca continues the general trend of the experimental systematics well, a spin–parity assignment of 3^- seems plausible for the 3,699-keV state. It is stressed, however, that these spin–parity quantum numbers could not be confirmed from the experimental data and are therefore suggested only as tentative assignments.

Shell-model predictions of excited states for ^{54}Ca are presented in Fig. 2b. Here we report calculations performed in the *sd*–*fp*–*sdg* model space (specifically, the $1d_{5/2}$, $1d_{3/2}$, $2s_{1/2}$, $1f_{7/2}$, $1f_{5/2}$, $2p_{3/2}$, $2p_{1/2}$, $1g_{9/2}$, $1g_{7/2}$, $2d_{5/2}$, $2d_{3/2}$ and $3s_{1/2}$ SPOs) using a modified GXPF1B shell-model Hamiltonian[14] with an adjustment to the strength (–0.15 MeV) of the $\nu p_{3/2}$–$\nu f_{5/2}$ monopole interaction between neutrons. Details on other components of the effective interaction are provided in ref. 26. The calculations indicate that the 2_1^+ state is primarily a consequence of a neutron particle–hole excitation across the $N = 34$ subshell gap and, despite $E(2_1^+)$ being lower than in ^{52}Ca as a result of the correlation energy, the strength of the $N = 34$ subshell gap (the $\nu p_{1/2}$–$\nu f_{5/2}$ SPO energy gap for ^{54}Ca) is in fact similar to the one at $N = 32$ (the $\nu p_{3/2}$–$\nu p_{1/2}$ SPO energy gap for ^{52}Ca). More specifically, our calculations suggest that although $E(2_1^+)$ for ^{52}Ca is very similar to the $\nu p_{3/2}$–$\nu p_{1/2}$ SPO energy gap at $N = 32$, the effect of the correlation energy reduces $E(2_1^+)$ by ~0.5 MeV relative to the $\nu p_{1/2}$–$\nu f_{5/2}$ SPO energy gap for ^{54}Ca. The difference is mainly attributed to the larger ground-state correlation energy of ^{52}Ca, which is understood from the relative strengths of the $\langle p_{3/2}p_{3/2}|V|p_{1/2}p_{1/2}\rangle_{J=0}$ and

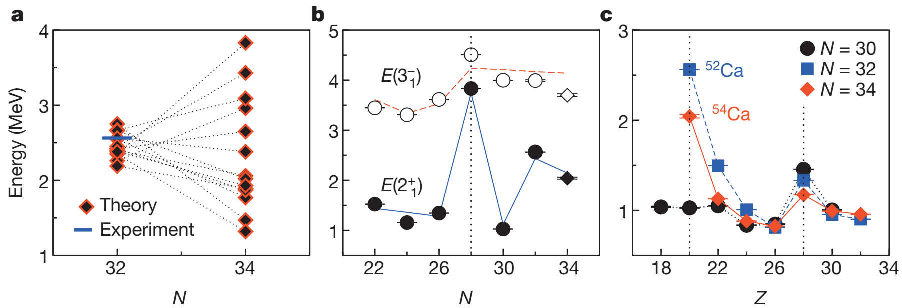


Figure 2 | Systematics of excited-state energies in even–even Ca isotopes and neighbouring nuclei. a, Theoretical predictions of the energy of the first 2^+ state for ^{52}Ca ($N = 32$) and ^{54}Ca ($N = 34$) (refs 14–16, 19–24). The solid blue line represents the experimental result for ^{52}Ca (refs 6, 7). **b**, Energies of the first 2^+ (filled symbols) and 3^- (open symbols) levels for even–even $^{42-54}$Ca isotopes. The results of the present study are indicated by diamonds at $N = 34$. The solid and dashed lines are shell-model predictions of the first 2^+ and 3^- energies, respectively (see text for details). **c**, $E(2_1^+)$ along the $N = 30$, 32 and 34 isotonic chains. The solid and dashed lines are intended to guide the eye. Vertical dotted lines represent the standard magic numbers.

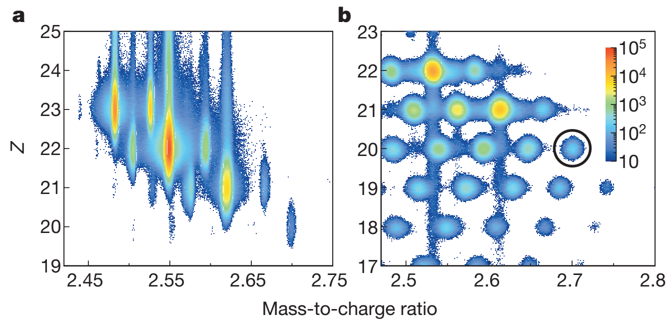


Figure 3 | Particle identification plots for radioisotopes. **a**, Constituents of the radioactive beam produced from the fragmentation of $^{70}Zn^{30+}$ ions and accepted by the BigRIPS separator. The most intense groups at $Z = 21$ and 22 are ^{55}Sc and ^{56}Ti, respectively. **b**, Reaction products accepted by the ZeroDegree spectrometer; ^{54}Ca events are enclosed by the black circle. The colour scale indicates the number of counts per histogram bin for both plots.

$\langle p_{1/2}p_{1/2}|V|f_{5/2}f_{5/2}\rangle_{J=0}$ pairing-matrix elements (-1.3 and -0.6 MeV, respectively). Thus, the first excited state of ^{54}Ca is understood to be lower in energy than its ^{52}Ca counterpart despite the sizes of the $N = 34$ and 32 subshell gaps being comparable.

The predicted 3_1^- state lies 0.4 MeV above its suggested experimental counterpart (the energy level at 3.7 MeV) and therefore provides a reasonable description of the data that supports the tentative spin–parity assignment. Although a $J^{\Pi} = 4^+$ assignment for the 3.7-MeV level cannot be completely ruled out, it is important to realize that in a previous investigation of ^{52}Ca, where the same general experimental techniques as those outlined here were used, direct population of the first 3^- state was reported from proton knockout reactions with no appreciable population of positive-parity excited states[7]. Because a similar population pattern is expected for ^{54}Ca in the present study, a spin–parity assignment of 3^- is favoured. Furthermore, our calculations and ref. 24 indicate that the first 4^+ state is expected to lie at ~4.5 MeV and is therefore unlikely to correspond to the 3.7-MeV experimental level.

In Fig. 2b, systematics of predicted $E(2_1^+)$ and $E(3_1^-)$ values from ^{42}Ca to ^{52}Ca are also displayed. The calculations satisfactorily reproduce the energy levels along the Ca isotopic chain from stability to exotic regions and therefore add further confidence to the spin–parity assignments discussed above. Moreover, we note that the modification to the GXPF1B Hamiltonian has no major effect on the predicted energy systematics along the Ti or Cr isotopic chains, thus preserving the success of the original shell-model interaction[14] for these nuclei. It is also noted that the theoretical studies of refs 19–21 provide satisfactory descriptions of the experimental $E(2_1^+)$ measurement for ^{54}Ca that further strengthen the spin–parity assignment.

More recently, calculations including three-body forces have been performed for Ca isotopes[23,24]. These mark the present frontier for such studies, and their importance has been discussed for new mass measurements of ^{53}Ca and ^{54}Ca (ref. 12). We note that the results of refs 23, 24 also provide reasonable descriptions of the experimental $E(2_1^+)$ value for ^{54}Ca reported here. The major relevant effect of three-body forces lies in the repulsive monopole interaction between valence neutrons, which is a consequence of the Fujita–Miyazawa force, as outlined in ref. 27. In the case of shell-model effective interactions, such as the modified GXPF1B Hamiltonian, where the π–ν tensor force drives structural evolution with varying Z (Fig. 1), the influence of three-body forces is included empirically and, hence, these models

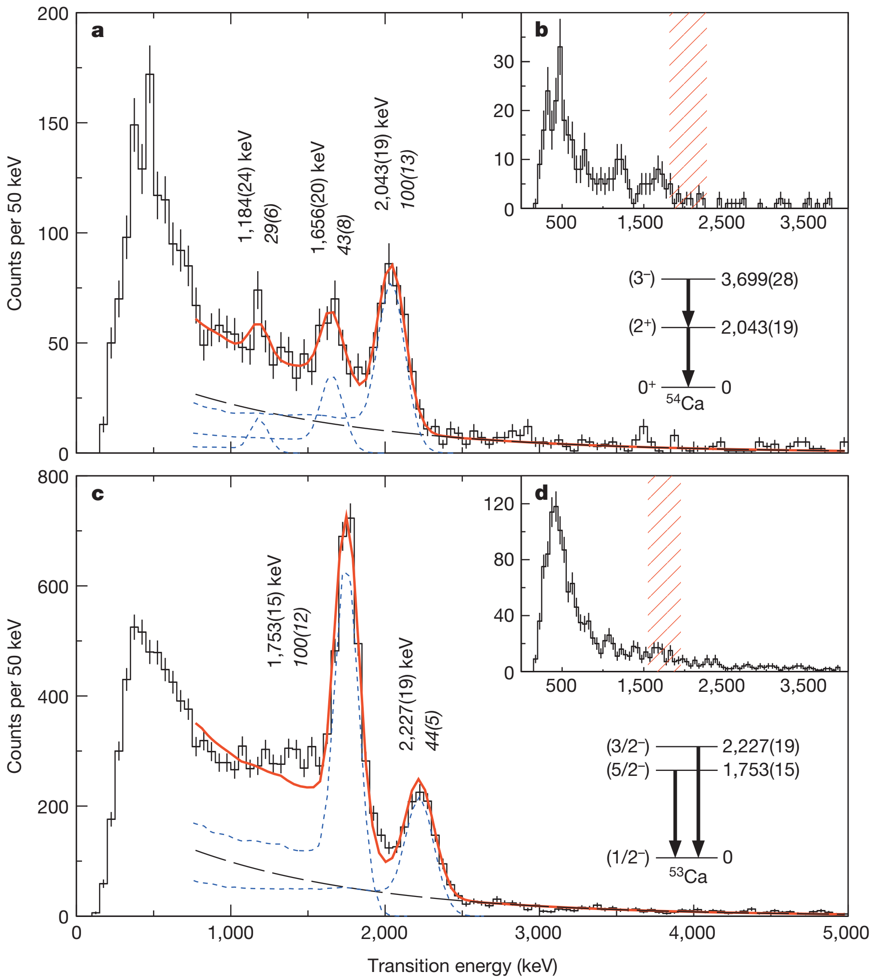


Figure 4 | Doppler-corrected γ-ray energy spectra. **a**, **c**, De-excitation γ-rays measured in coincidence with ^{54}Ca (**a**) and ^{53}Ca (**c**) reaction products. Peaks are labelled by their energies and γ-ray relative intensities are indicated by italic text. The short blue and long black dashed lines represent simulations performed with the code GEANT4[30] and exponential fits to the background components, respectively. Experimental level schemes are presented in the lower-right sections, with tentative spin–parity assignments in parentheses. **b**, **d**, ^{54}Ca (**b**) and ^{53}Ca (**d**) γ-rays measured in coincidence with the 2,043- and 1,753-keV transitions, respectively. The red-hatched areas represent the widths of the energy gates used in the coincidence measurements. Error bars displayed on all histogram bins are the square roots of the bin values. The errors quoted for γ-ray energies, γ-ray relative intensities and level energies represent systematic and statistical (1 s.d.) uncertainties added in quadrature.

can produce results similar to those of theories that treat three-body forces more explicitly.

In addition to the energy levels of ^{54}Ca, the structure of ^{53}Ca may also provide key information that reflects the strength of the $N = 34$ subshell closure. Figure 4c displays a γ-ray energy spectrum for ^{53}Ca deduced in the present study that indicates two peaks at 1,753(15) and 2,227(19) keV. The 1,753-keV line has not been reported before, but the transition at 2,227 keV is consistent in energy with a peak at 2,220(1) keV previously reported in a decay study[28]. The γγ coincidence measurements presented in the spectrum of Fig. 4d indicate that the two transitions are not emitted in the same decay sequence; accordingly, the γ-rays are placed in parallel in the ^{53}Ca level scheme (Fig. 4c), where two excited states at 1,753(15) and, respectively, 2,227(19) keV decay directly to the ground state. The result of ref. 28 and the relative intensities of the two peaks deduced in the present work confirm the proposed level scheme.

The spins and parities of the ^{53}Ca excited states could not be deduced from the experimental data. The spin–parity assignment of $1/2^-$ for the ^{53}Ca ground state is adopted from simple shell-model arguments, because the unpaired neutron is expected to occupy the $\nu p_{1/2}$ SPO. This is consistent with the assignments in refs 18, 28. Tentative spin–parity quantum numbers of $5/2^-$ and $3/2^-$ have been assigned to the 1,753- and 2,227-keV levels on the basis of the decay study[28] and predictions of the shell model using the modified GXPF1B Hamiltonian, which indicates $5/2^-$ and $3/2^-$ excited states at 1,930 and 2,075 keV, respectively. These states are understood to be neutron excitations across the respective $N = 34$ ($\nu p_{1/2} \rightarrow \nu f_{5/2}$) and $N = 32$ ($\nu p_{3/2} \rightarrow \nu p_{1/2}$) subshell gaps. We note that the theoretical studies of refs 20, 21, 24 support the proposed ordering of the spin–parity assignments.

Our results provide direct experimental evidence for a new nuclear 'magic number' at $N = 34$ in exotic calcium isotopes, the existence of which has been a subject of much debate over the past decade[13], and further emphasize the evolution of nuclear shell structure in systems far from stability. In addition, we have shown that the energy of the 2_1^+ state of ^{54}Ca, measured here by in-beam γ-ray spectroscopy, is essential for clarifying sizable discrepancies among theoretical results. From a more general standpoint, increasing our knowledge of the structures of exotic nuclei and the nucleon–nucleon forces that govern nuclear shell evolution in systems far from stability has a key role in improving our understanding of astrophysical processes such as nucleosynthesis. Finally, we note that future measurements of excited states and masses for the neighbouring isotopes ^{55}Ca and ^{56}Ca are encouraged, because they now represent the experimental frontier for Ca and will further enhance our understanding of nuclear quantum states in very neutron-rich systems.

METHODS SUMMARY

A radioactive beam partly composed of the $N = 34$ isotones ^{55}Sc and ^{56}Ti was produced by projectile fragmentation of ^{70}Zn^{30+} ions at 345 MeV per nucleon. The radioactive beam, with constituents travelling at approximately 60% of the speed of light, was selected according to particle magnetic rigidities (momentum-to-charge ratios) with the BigRIPS separator[25], and hardware slits were used to reduce beam contamination. The magnetic spectrometer was optimized for the transportation of ^{55}Sc. The constituents of the radioactive beam were identified on an event-by-event basis using standard techniques in which the mass-to-charge ratio and proton number of each projectile were deduced. Details of such methods are provided, for example, in ref. 29. A 10-mm-thick Be reaction target was inserted in a high-efficiency γ-ray detector array consisting of 186 NaI(Tl) scintillator crystals with almost full solid-angle (4π sr) coverage. The array was calibrated with respect to γ-ray energy using standard ^{60}Co, ^{88}Y and ^{137}Cs sources. Reaction products were identified with the ZeroDegree spectrometer[25], which was tuned for the transmission of ^{54}Ca, using the same general techniques as for the BigRIPS separator. The maximum intensity of the ^{70}Zn primary beam was ~100 particle nanoamperes, which corresponds to 6.3×10^{11} particles per second (p.p.s.), and the rates of ^{55}Sc and ^{56}Ti delivered to the reaction target were 12 and 125 p.p.s. per particle nanoampere of primary beam, respectively. Data were accumulated over a time period of approximately two days with the condition of requiring a particle in the ZeroDegree spectrometer in coincidence with at least one γ-ray. A total of 3.8×10^4 ^{54}Ca-γ^n events (γ-ray multiplicity, $n \geq 1$) were collected at an average rate of 1,050 per hour.

Received 16 May; accepted 1 August 2013.

1. Goeppert Mayer, M. On closed shells in nuclei. II. *Phys. Rev.* **75,** 1969–1970 (1949).
2. Warner, D. Not-so-magic numbers. *Nature* **430,** 517–519 (2004).
3. Janssens, R. V. F. Unexpected doubly magic nucleus. *Nature* **459,** 1069–1070 (2009).
4. Bastin, B. *et al.* Collapse of the $N = 28$ shell closure in ^{42}Si. *Phys. Rev. Lett.* **99,** 022503 (2007).
5. Otsuka, T., Suzuki, T., Fujimoto, R., Grawe, H. & Akaishi, Y. Evolution of nuclear shells due to the tensor force. *Phys. Rev. Lett.* **95,** 232502 (2005).
6. Huck, A. *et al.* Beta decay of the new isotopes ^{52}K, ^{52}Ca, and ^{52}Sc; a test of the shell model far from stability. *Phys. Rev. C* **31,** 2226–2237 (1985).
7. Gade, A. *et al.* Cross-shell excitation in two-proton knockout: structure of ^{52}Ca. *Phys. Rev. C* **74,** 021302(R) (2006).
8. Janssens, R. V. F. *et al.* Structure of ^{52,54}Ti and shell closures in neutron-rich nuclei above ^{48}Ca. *Phys. Lett. B* **546,** 55–62 (2002).
9. Dinca, D.-C. *et al.* Reduced transition probabilities to the first 2^+ state in ^{52,54,56}Ti and development of shell closures at $N = 32, 34$. *Phys. Rev. C* **71,** 041302(R) (2005).
10. Chapman, R., Hinds, S. & MacGregor, A. E. A study of ^{52}Cr, ^{54}Cr and ^{56}Cr by the (t,p) reaction. *Nucl. Phys. A* **119,** 305–324 (1968).
11. Bürger, A. *et al.* Relativistic Coulomb excitation of neutron-rich ^{54,56,58}Cr: on the pathway of magicity from $N = 40$ to $N = 32$. *Phys. Lett. B* **622,** 29–34 (2005).
12. Wienholtz, F. *et al.* Masses of exotic calcium isotopes pin down nuclear forces. *Nature* **498,** 346–349 (2013).
13. Otsuka, T. *et al.* Magic numbers in exotic nuclei and spin-isospin properties of the NN interaction. *Phys. Rev. Lett.* **87,** 082502 (2001).
14. Honma, M., Otsuka, T. & Mizusaki, T. Shell-model description of neutron-rich Ca isotopes. *RIKEN Accel. Prog. Rep.* **41,** 32 (2008).
15. Liddick, S. N. *et al.* Lowest excitations in ^{56}Ti and the predicted $N = 34$ shell closure. *Phys. Rev. Lett.* **92,** 072502 (2004).
16. Prisciandaro, J. I. *et al.* New evidence for a subshell gap at $N = 32$. *Phys. Lett. B* **510,** 17–23 (2001).
17. Janssens, R. V. F. Elusive magic numbers. *Nature* **435,** 897–898 (2005).
18. Crawford, H. L. *et al.* β decay and isomeric properties of neutron-rich Ca and Sc isotopes. *Phys. Rev. C* **82,** 014311 (2010).
19. Rodríguez, T. R. & Egido, J. L. New beyond-mean-field theories: examination of the potential shell closures at $N = 32$ or 34. *Phys. Rev. Lett.* **99,** 062501 (2007).
20. Rejmund, M. *et al.* Shell evolution and the $N = 34$ "magic number". *Phys. Rev. C* **76,** 021304(R) (2007).
21. Coraggio, L., Covello, A., Gargano, A. & Itaco, N. Spectroscopic study of neutron-rich calcium isotopes with a realistic shell-model interaction. *Phys. Rev. C* **80,** 044311 (2009).
22. Kaneko, K., Sun, Y., Mizusaki, T. & Hasegawa, M. Shell-model study for neutron-rich *sd*-shell nuclei. *Phys. Rev. C* **83,** 014320 (2011).
23. Holt, J. D., Otsuka, T., Schwenk, A. & Suzuki, T. Three-body forces and shell structure in calcium isotopes. *J. Phys. G* **39,** 085111 (2012).
24. Hagen, G., Hjorth-Jensen, M., Jansen, G. R., Machleidt, R. & Papenbrock, T. Evolution of shell structure in neutron-rich calcium isotopes. *Phys. Rev. Lett.* **109,** 032502 (2012).
25. Kubo, T. *et al.* BigRIPS separator and ZeroDegree spectrometer at RIKEN RI Beam Factory. *Prog. Theor. Exp. Phys.* 03C003 (2012).
26. Utsuno, Y. *et al.* Shell evolution around and beyond $N = 28$ studied with large-scale shell-model calculations. *Prog. Theor. Phys.* **196** (suppl.), 304–309 (2012).
27. Otsuka, T., Suzuki, T., Holt, J. D., Schwenk, A. & Akaishi, Y. Three-body forces and the limit of oxygen isotopes. *Phys. Rev. Lett.* **105,** 032501 (2010).
28. Perrot, F. *et al.* β-decay studies of neutron-rich K isotopes. *Phys. Rev. C* **74,** 014313 (2006).
29. Ohnishi, T. *et al.* Identification of new isotopes ^{125}Pd and ^{126}Pd produced by in-flight fission of 345 MeV/nucleon ^{238}U: first results from the RIKEN RI Beam Factory. *J. Phys. Soc. Jpn* **77,** 083201 (2008).
30. Agostinelli, S. *et al.* GEANT4—a simulation toolkit. *Nucl. Instrum. Methods A* **506,** 250–303 (2003).

Acknowledgements We thank the staff of the RIKEN Nishina Center accelerator complex for their contribution to the experiment. This work is part of the CNS-RIKEN joint research project on large-scale nuclear structure calculations.

Author Contributions D.S. performed offline data analyses and GEANT4 simulations, wrote the manuscript and prepared the figures. D.S. and S.T. designed the experiment. N.A., P.D., J.L., K.M., M.M., Y.S., D.S., S.T., R.T. and H.W. were responsible for setting up the γ-ray detector array. D.N. prepared the beamline ionization chambers. N.F. and H. Suzuki tuned the BigRIPS separator and the ZeroDegree spectrometer. H.B. was responsible for the data acquisition system. N.A., P.D., S.G., J.L., K.M., M.M., S.M., Y.S., P.-A.S., D.S., T.S., S.T., R.T., J.J.V.-D., H.W. and K.Y. checked data accumulation online and maintained operation of the experiment. M.H., T.O. and Y.U. performed the shell-model calculations. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.S. (steppenbeck@cns.s.u-tokyo.ac.jp).

LETTER

doi:10.1038/nature12539

Observing single quantum trajectories of a superconducting quantum bit

K. W. Murch[1,2], S. J. Weber[1], C. Macklin[1] & I. Siddiqi[1]

The length of time that a quantum system can exist in a superposition state is determined by how strongly it interacts with its environment. This interaction entangles the quantum state with the inherent fluctuations of the environment. If these fluctuations are not measured, the environment can be viewed as a source of noise, causing random evolution of the quantum system from an initially pure state into a statistical mixture—a process known as decoherence. However, by accurately measuring the environment in real time, the quantum system can be maintained in a pure state and its time evolution described by a 'quantum trajectory'[1,2] determined by the measurement outcome. Here we use weak measurements to monitor a microwave cavity containing a superconducting quantum bit (qubit), and track the individual quantum trajectories[3] of the system. In this set-up, the environment is dominated by the fluctuations of a single electromagnetic mode of the cavity. Using a near-quantum-limited parametric amplifier[4,5], we selectively measure either the phase or the amplitude of the cavity field, and thereby confine trajectories to either the equator or a meridian of the Bloch sphere. We perform quantum state tomography at discrete times along the trajectory to verify that we have faithfully tracked the state of the quantum system as it diffuses on the surface of the Bloch sphere. Our results demonstrate that decoherence can be mitigated by environmental monitoring, and validate the foundation of quantum feedback approaches based on Bayesian statistics[6–8]. Moreover, our experiments suggest a new means of implementing 'quantum steering'[9]—the harnessing of action at a distance to manipulate quantum states through measurement.

If a quantum system and its environment have a common set of stationary states, then a measurement of the environment ultimately leads to projection of the quantum system onto one of its eigenstates. Numerous experiments with photons[10], atoms[11] and solid-state systems[12,13] have elucidated this process, enabling squeezing[14–16] and quantum feedback[6–8,17]. Here we examine how real-time monitoring of a quantum system's environment reveals quantum trajectories that underlie the processes of measurement and decoherence.

In our experiment (Fig. 1a), we use a superconducting transmon circuit[18] dispersively coupled to a copper waveguide cavity in the 'three-dimensional transmon' architecture[19]. If only the two lowest levels of the transmon are considered, the qubit–cavity interaction is given by the Hamiltonian $H_{\mathrm{int}} = -\hbar\chi a^{\dagger}a\sigma_z$, where $a^{\dagger}$ and a are respectively the creation and annihilation operators for the cavity mode, σ_z is the qubit Pauli operator that acts on the qubit state in the energy basis, χ is the dispersive coupling rate and $\hbar$ is Planck's constant divided by 2π. This interaction can be viewed as either a qubit-state-dependent shift of the cavity frequency of $-\chi\sigma_z$ or a light (or a.c. Stark) shift of the qubit

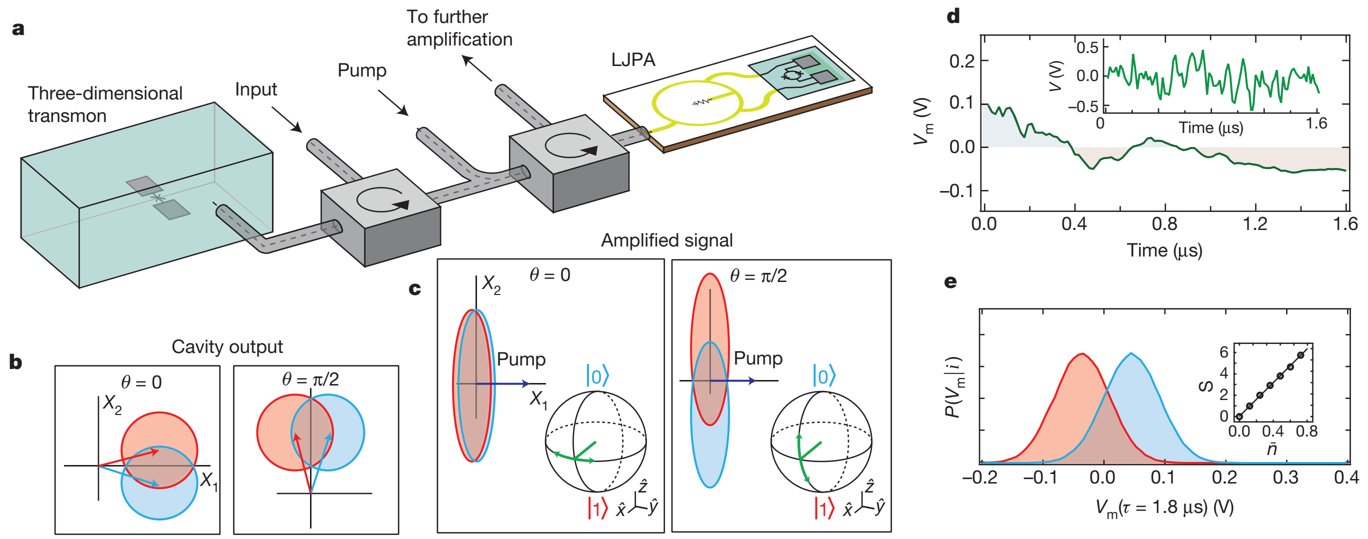


Figure 1 | Single-quadrature weak measurements. **a**, Our experimental set-up consists of a superconducting transmon qubit dispersively coupled to a copper waveguide cavity with a coupling rate of $\chi/2\pi = -0.49$ MHz. The cavity port sets the cavity decay rate of $\kappa/2\pi = 10.8$ MHz. Signals that reflect from the cavity are amplified by a LJPA (lumped-element Josephson parametric amplifier) operating with 10 dB of gain and an instantaneous bandwidth of 20 MHz. **b**, Illustration of the Gaussian variance of measurement signals with phases, relative to the amplifier pump, of $\theta = 0$ and $\pi/2$ after reflecting from the cavity with the qubit in $|1\rangle$ (red) or $|0\rangle$ (blue). **c**, Illustration of the measurement signal after reflecting from the LJPA. The amplifier is operated in phase-sensitive mode, where small signals that are combined in phase or in quadrature with the pump tone are respectively amplified or de-amplified, and rotated by $\pi/2$. The back-action of the measurement on a qubit superposition state is indicated on the Bloch sphere. **d**, A representative integrated measurement signal, $V_{\mathrm{m}}(\tau)$, that is obtained when the qubit is prepared in an initial superposition state along the $\hat{x}$ axis of the Bloch sphere. Inset, instantaneous measurement voltage. **e**, Histograms of $V_{\mathrm{m}}(\tau = 1.8\,\mu\mathrm{s})$ for the qubit prepared in $|1\rangle$ (red) and $|0\rangle$ (blue) for a measurement corresponding to $\theta = \pi/2$ with $S = 3.2$. Inset, S versus $\bar{n}$, with the solid line indicating the expected dependence for $\eta = 0.49$.

[1]Quantum Nanoelectronics Laboratory, Department of Physics, University of California, Berkeley, California 94720, USA. [2]Department of Physics, Washington University, St Louis, Missouri 63130, USA.

frequency that depends on the intracavity photon number $\hat{n} = a^{\dagger}a$. A microwave tone that probes the cavity near its resonance frequency acquires a qubit-state-dependent phase shift (Fig. 1b). For $|\chi| \ll \kappa$, where κ is the cavity decay rate, the reflected signal component in quadrature with the input measurement tone contains qubit state information, and the signal component in phase with the measurement tone carries information about the intracavity photon number. After leaving the cavity, the signal is displaced to the origin of the X_1–X_2 plane (X_1 and X_2 are the two quadrature amplitudes of the mode of the electromagnetic field) by a coherent tone and amplified by a near-quantum-limited lumped-element Josephson parametric amplifier[5] (LJPA). Phase-sensitive operation of the LJPA permits noiseless amplification of one quadrature of the reflected measurement signal with corresponding de-amplification of the other quadrature[20,21].

In principle, for an ideal amplifier, that is, one which adds no noise, the choice of measurement quadrature determines the type of back-action imparted on a coherent superposition of qubit states[22]. When the amplified quadrature conveys information about the qubit state, the measurement causes random motions of the qubit towards its eigenstates, which are located at the poles of the Bloch sphere. From the perspective of the qubit, the intracavity photon number does not fluctuate. However, when the amplified quadrature encodes the intracavity photon number, the phase of the coherent superposition evolves in response to variations in the a.c. Stark shift of the qubit and superpositions of the measurement eigenstates are not projected.

We first focus on the case in which the amplified quadrature conveys qubit state information, which we denote a 'Z-measurement'. Figure 1d displays a single measurement signal, $V_{\mathrm{m}}(\tau) = (1/\tau)\int_0^{\tau} V(t)\,\mathrm{d}t$, where $V(t)$ is the instantaneous measurement voltage, that is obtained when the qubit is initialized in the superposition state $(|0\rangle + |1\rangle)/\sqrt{2}$ along the $\hat{x}$ axis of the Bloch sphere. As the measurement duration increases, information about the qubit state accumulates. The best estimate for the state of the qubit after a weak measurement can be obtained using Bayes' rule[10,22,23], to find the probability of the qubit being in state i, conditioned on the integrated measurement value V_{m}:

$$P(i|V_{\mathrm{m}}) = \frac{P(i)P(V_{\mathrm{m}}|i)}{P(V_{\mathrm{m}})}$$

Probability distributions $P(V_{\mathrm{m}}|i)$ for V_{m} are shown in Fig. 1e for the states $i = \{|0\rangle, |1\rangle\}$. The probability $P(i)$ describes our knowledge of the prior distribution and is 1/2 when the qubit is initialized along $\hat{x}$. Thus, after acquiring a measurement value V_{m}, the state of the system is described by[22]

$$Z^Z = \tanh(V_{\mathrm{m}}S/2\Delta V) \tag{1}$$

$$X^Z = \sqrt{1 - (Z^Z)^2}\,\mathrm{e}^{-\gamma\tau} \tag{2}$$

Here we define the expectation values of the Pauli operators conditioned on measurement value V_{m} as $X = \langle \sigma_x \rangle|_{V_{\mathrm{m}}}$, $Y = \langle \sigma_y \rangle|_{V_{\mathrm{m}}}$ and $Z = \langle \sigma_z \rangle|_{V_{\mathrm{m}}}$, where the superscript 'z' denotes a Z-measurement and $S = 64\tau\chi^2\bar{n}\eta/\kappa$ is the dimensionless measurement strength that depends on the measurement duration, τ, the quantum efficiency of the measurement, η, and the average intracavity photon number, $\bar{n}$. The measurement strength can also be related to the separation between the measurement probability distributions for the states $|0\rangle$ and $|1\rangle$, ΔV, and their Gaussian variance, σ^2: $S = (\Delta V)^2/\sigma^2$. We calibrate $\bar{n}$ using the measured a.c. Stark shift of the qubit frequency. From a linear fit of S to $\bar{n}$ (Fig. 1e, inset), we determine that $\eta = 0.49$. For small values of S, an individual measurement does not fully determine the qubit state. For $S \gg 1$, the histograms are well separated and the qubit state can be determined with very high confidence, corresponding to a projective measurement.

The exponential decay of coherence in X^Z in equation (2) reflects imperfect knowledge about the state of the environment and leads to qubit dephasing characterized by the rate $\gamma = 8\chi^2\bar{n}(1-\eta)/\kappa + 1/T_2^*$. The first term here reflects measurement-induced dephasing[24,25] originating from the degree to which the signal is undetected (quantified by $1-\eta$). The second term reflects extra environmental dephasing, which is characterized by $T_2^* = 20\ \mu\mathrm{s}$ for the qubit.

We now discuss the case of a 'ϕ-measurement', in which the amplified quadrature encodes the fluctuating intracavity photon number. Each photon that enters the cavity shifts the qubit phase by an average of $4\chi/\kappa$, causing the phase of a coherent superposition of the qubit states $|0\rangle$ and $|1\rangle$ to evolve. Given ΔV and S (obtained from a separate Z-measurement), V_{m} can be used to infer the total accrued phase shift. The evolution of X and Y is then given by[22]

$$X^{\phi} = \cos(SV_{\mathrm{m}}/2\Delta V)\mathrm{e}^{-\gamma\tau} \tag{3}$$

$$Y^{\phi} = -\sin(SV_{\mathrm{m}}/2\Delta V)\mathrm{e}^{-\gamma\tau} \tag{4}$$

The back-action associated with quadrature specific amplification, as given by equations (1)–(4), is presented in Fig. 2. To verify these predictions, we conduct an experiment consisting of three primary actions: we first prepare the qubit along the $\hat{x}$ axis, then we digitize the amplified measurement tone for 1.8 µs and, finally, we measure the projection of the qubit state along the $\hat{x}$, $\hat{y}$ or $\hat{z}$ axis. After repeating the experiment sequence $\sim 10^5$ times, we evaluate X^Z, X^{ϕ}, Y^Z, Y^{ϕ}, Z^Z and Z^{ϕ}. Figure 2c displays the results of this measurement procedure for a Z-measurement, as a function of V_{m}. A measurement with $V_{\mathrm{m}} = 0$ yields no information about the qubit state, leaving it unperturbed. A strongly positive or negative value of V_{m}, although rare, corresponds to a significant motion of the qubit towards the state $|0\rangle$ or, respectively, the state $|1\rangle$. For the ϕ-measurement (Fig. 2d), Z^{ϕ} is uncorrelated with the measurement signal. Here V_{m} conveys information about the phase shift of the qubit state resulting from the fluctuating intracavity photon number.

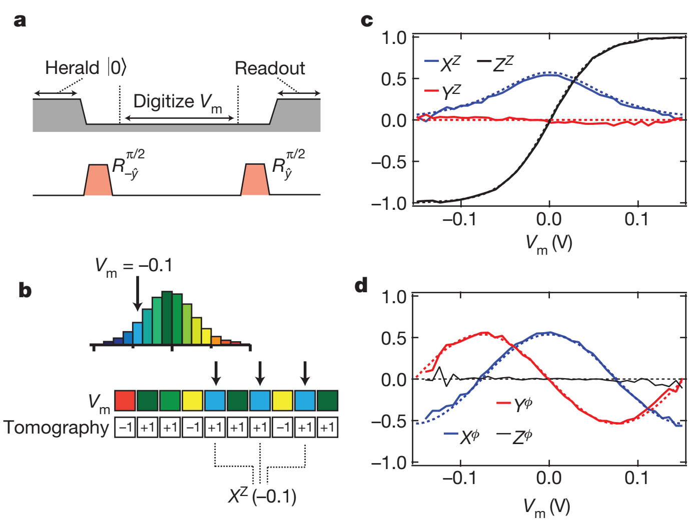


Figure 2 | Correlation of tomography results with measurement values. **a**, Experimental sequence for determining X^Z. An initial strong measurement is used to herald $|0\rangle$, and this is followed by a rotation about the $\hat{y}$ axis to prepare the qubit along $\hat{x}$. A weaker measurement signal is digitized for 1.8 µs, and a final rotation and strong measurement are used to determine the qubit projection along $\hat{x}$. Similar sequences are used to determine Y^Z and Z^Z. **b**, Tomography correlation procedure. Different measurement values are indicated as different colours, with the colour-coded histogram indicating the relative probability of each measurement value. Boxes indicate the measurement value, V_{m}, for each experimental repetition, and the associated tomography result is indicated as ± 1. Tomography results for matching V_{m} are averaged together to determine X^Z. **c**, Tomography results versus V_{m} for a Z-measurement with $\bar{n} = 0.4$. The dashed lines are theory curves based on equations (1) and (2) for $\eta = 0.49$ and $S = 3.15$, where $\gamma = 2.8 \times 10^5\ \mathrm{s}^{-1}$. **d**, Tomography results for a ϕ-measurement for $\bar{n} = 0.46$. The dashed lines are theory curves based on equations (3) and (4) for $\eta = 0.49$ and $S = 3.62$, where $\gamma = 3.1 \times 10^5\ \mathrm{s}^{-1}$.

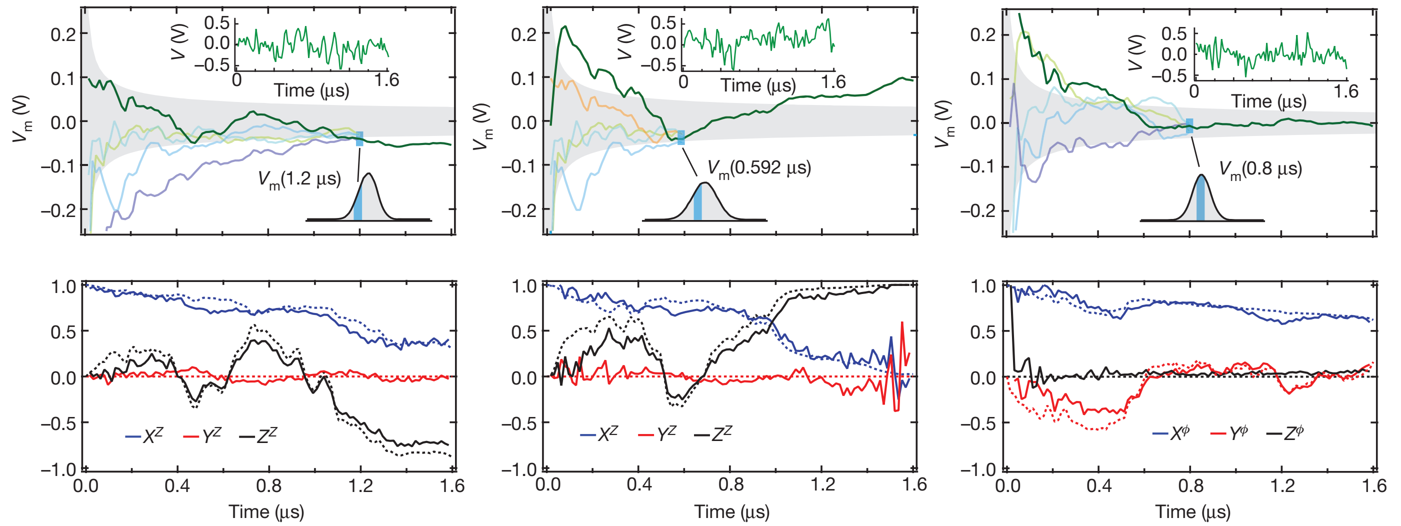


Figure 3 | Quantum trajectories. a, b, Individual measurement traces obtained for Z-measurements with $\bar{n}=0.4$. The top panels display $V_m(\tau)$ as a green line, with the upper insets displaying the instantaneous measurement voltage. The grey region indicates the standard deviation of the distribution of measurement values. Measurement traces that converge to an integrated value within the blue matching window are used to reconstruct, using tomography, the trajectory at that time point. A few different measurement traces that contribute to the reconstruction at 1.2 µs (**a**) and 0.592 µs (**b**) are indicated in lighter colours. The lower insets indicate the distribution of measurement values with the blue matching window. The lower panels display quantum trajectories obtained from analysis of the measurement signal, as dotted lines. Solid lines indicate the tomographically reconstructed quantum trajectory based on the ensemble of measurements that are within the matching window of the original measurement signal. **c**, Individual measurement traces and associated quantum trajectories obtained for a ϕ-measurement with $\bar{n}=0.4$.

For both types of measurement, the results show excellent agreement with theory for $\eta = 0.49$.

We have so far demonstrated that the integrated measurement signal provides a faithful record of the fluctuations of the environment and the associated motions of the qubit state. Moreover, we observe that the direction of motion of the qubit state depends on the amplification quadrature. To examine quantum trajectories of the system, we divide the measurement signal into successive segments. The integrated measurement signal can then be written as a string $\{V_m(\tau_0), V_m(\tau_1), V_m(\tau_2), \ldots\}$, where $\tau_{i+1} - \tau_i = 16$ ns. At each time point, $V_m(\tau_i)$ can be used to infer the qubit state. In Fig. 3, we present measurement traces along with the quantum trajectory of the system associated with each noisy measurement trace. The trajectories show how the quantum system evolves stochastically from an initial state prepared along $\hat{x}$ towards a final state. Measurement inefficiency and additional dephasing limits the accuracy with which the state can be tracked. This limitation is manifest as a gradual shortening of the estimated transverse coherence of the qubit state.

To verify that we have accurately inferred the quantum trajectory of the system corresponding to a given measurement signal, we perform quantum state tomography on an ensemble of experimental iterations with similar measurement values. A tomographic reconstruction of the trajectory is obtained by making measurements of variable duration, τ_i, and subsequently measuring the projection of the qubit state along one of the Cartesian axes of the Bloch sphere. Only measurements with values that are within $\pm\varepsilon$ of the target value, $V_m(\tau_i)$, contribute to determining the ensemble properties X, Y and Z. As shown in the upper panels of Fig. 3, many different measurement signals that converge to $V_m(\tau_i)\pm\varepsilon$ at τ_i are used in the tomographic reconstruction. This illustrates how the inferred state at a particular time depends only on the value of the integrated measurement voltage at that time.

Figure 3a, b displays quantum trajectories that are obtained for Z-measurements. The reconstructed trajectories based on ensemble measurements, shown as solid lines, are in reasonable agreement with the quantum trajectories determined from a single measurement record, and reproduce many of the minute motions of the qubit as it ultimately evolves towards its eigenstates of measurement. Some trajectories highlight the concept of quantum measurement reversal:[26–29] in Fig. 3a, after ~400, 600 and 1,000 ns of measurement the qubit state has returned nearly to its original state, effectively 'reversing' the preceding partial collapse of the qubit wavefunction. In Fig. 3c, we display the measurement record that we obtain from a ϕ-measurement. The resulting quantum trajectory is confined to motions along the equator of the Bloch sphere.

Full control over the environment of a quantum system allows for the mitigation of decoherence through accurate monitoring of fluctuations of the environment. Although measurement schemes based on projective measurements on ancilla qubits obtain measurement efficiencies[30] greater than 0.9, the measurement efficiency presented here, $\eta = 0.49$, is among the highest reported values for a continuous variable[6–8,13]. This efficiency is limited by an imperfect collection efficiency, $\eta_{col} = 0.72$, resulting from losses in microwave components, and imperfect amplifier quantum efficiency, $\eta_{amp} = 0.68$. Further improvements in the quantum measurement efficiency will be essential for realizing potential applications of quantum feedback[6,7] in quantum metrology and information science.

METHODS SUMMARY

The qubit consists of two aluminium paddles connected by a double-angle-evaporated aluminium superconducting quantum interference device (SQUID) deposited on double-side-polished silicon, and is characterized by charging and Josephson energies $E_c/h = 200$ MHz and $E_J/h = 11$ GHz, respectively. The qubit is operated with negligible flux threading the SQUID loop with transition frequency $\omega_q/2\pi = 3.999$ GHz. The qubit is located off centre of a 6.8316-GHz copper waveguide cavity.

The LJPA consists of a two-junction SQUID, formed from 2-µA Josephson junctions, shunted by 3 pF of capacitance, and is flux-biased to provide 10 dB of gain at the cavity resonance frequency. The LJPA is pumped by two sidebands that are equally spaced 300 MHz above and below the cavity resonance. A second LJPA that follows the first provides additional gain. A detailed experimental schematic is shown in Supplementary Fig. 1.

Experiment sequences start with an 800-ns readout with $S = 42$ that is used to herald the state $|0\rangle$ at the beginning of the experiment. A sample herald histogram is shown in Supplementary Fig. 2. Because $|\chi| \ll \kappa$, several peaks are visible, corresponding to the many energy levels of the transmon qubit. After preparing the state $|0\rangle$, we perform a 16-ns $\pi/2$-rotation about the $-\hat{y}$ axis to initialize the qubit along the $\hat{x}$ axis. After a period of variable duration, we perform quantum state tomography, which consists of either rotations about the $\hat{x}$ and $\hat{y}$ axes or no rotation and a second 800-ns readout with $S = 42$. In a fraction (~4%) of the final readouts, the qubit is outside the $\{|0\rangle, |1\rangle\}$ manifold. These sequences were disregarded in the analysis. Tomography results are corrected for the readout fidelity of 95%.

Received 29 May; accepted 7 August 2013.

1. Carmichael, H. *An Open Systems Approach to Quantum Optics* (Springer, 1993).
2. Carmichael, H. J., Kochan, P. & Tian, L. In *Proc. Int. Symp. Coherent States: Past, Present, and Future* (eds Feng, D. H., Strayer, M. R. & Klauder, J. R.) 75–91 (World Scientific, 1994).
3. Gambetta, J. *et al.* Quantum trajectory approach to circuit QED: quantum jumps and the Zeno effect. *Phys. Rev. A* **77,** 012112 (2008).
4. Castellanos-Beltran, M. A., Irwin, K. D., Hilton, G. C., Vale, L. R. & Lehnert, K. W. Amplification and squeezing of quantum noise with a tunable Josephson metamaterial. *Nature Phys.* **4,** 929–931 (2008).
5. Hatridge, M., Vijay, R., Slichter, D. H., Clarke, J. & Siddiqi, I. Dispersive magnetometry with a quantum limited squid parametric amplifier. *Phys. Rev. B* **83,** 134501 (2011).
6. Sayrin, C. *et al.* Real-time quantum feedback prepares and stabilizes photon number states. *Nature* **477,** 73–77 (2011).
7. Vijay, R. *et al.* Stabilizing Rabi oscillations in a superconducting qubit using quantum feedback. *Nature* **490,** 77–80 (2012).
8. Campagne-Ibarcq, P. *et al.* Persistent control of a superconducting qubit by stroboscopic measurement feedback. *Phys. Rev. X* **3,** 021008 (2013).
9. Schrödinger, E. Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften* **23,** 807–812,–823–824, 844–849 (1935).
10. Guerlin, C. *et al.* Progressive field-state collapse and quantum non-demolition photon counting. *Nature* **448,** 889–893 (2007).
11. Kuzmich, A. *et al.* Quantum nondemolition measurements of collective atomic spin. *Phys. Rev. A* **60,** 2346–2350 (1999).
12. Vijay, R., Slichter, D. H. & Siddiqi, I. Observation of quantum jumps in a superconducting artificial atom. *Phys. Rev. Lett.* **106,** 110502 (2011).
13. Hatridge, M. *et al.* Quantum back-action of an individual variable-strength measurement. *Science* **339,** 178–181 (2013).
14. Kuzmich, A., Mandel, L. & Bigelow, N. P. Generation of spin squeezing via continuous quantum nondemolition measurement. *Phys. Rev. Lett.* **85,** 1594–1597 (2000).
15. Takano, T., Fuyama, M., Namiki, R. & Takahashi, Y. Spin squeezing of a cold atomic ensemble with the nuclear spin of one-half. *Phys. Rev. Lett.* **102,** 033601 (2009).
16. Schleier-Smith, M. H., Leroux, I. D. & Vuletić, V. States of an ensemble of two-level atoms with reduced quantum uncertainty. *Phys. Rev. Lett.* **104,** 073604 (2010).
17. Ristè, D., Bultink, C. C., Lehnert, K. W. & DiCarlo, L. Feedback control of a solid-state qubit using high-fidelity projective measurement. *Phys. Rev. Lett.* **109,** 240502 (2012).
18. Koch, J. *et al.* Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **76,** 042319 (2007).
19. Paik, H. *et al.* Observation of high coherence in Josephson junction qubits measured in a three-dimensional circuit QED architecture. *Phys. Rev. Lett.* **107,** 240501 (2011).
20. Caves, C. M. Quantum limits on noise in linear amplifiers. *Phys. Rev. D* **26,** 1817–1839 (1982).
21. Clerk, A. A., Devoret, M. H., Girvin, S. M., Marquardt, F. & Schoelkopf, R. J. Introduction to quantum noise, measurement, and amplification. *Rev. Mod. Phys.* **82,** 1155–1208 (2010).
22. Korotkov, A. N. Quantum Bayesian approach to circuit QED measurement. Preprint at http://arxiv.org/abs/1111.4016 (2011).
23. Wiseman, H. & Milburn, G. *Quantum Measurement and Control* (Cambridge Univ. Press, 2010).
24. Schuster, D. I. *et al.* ac Stark shift and dephasing of a superconducting qubit strongly coupled to a cavity field. *Phys. Rev. Lett.* **94,** 123602 (2005).
25. Boissonneault, M., Gambetta, J. M. & Blais, A. Dispersive regime of circuit QED: photon-dependent qubit dephasing and relaxation rates. *Phys. Rev. A* **79,** 013819 (2009).
26. Korotkov, A. N. & Jordan, A. N. Undoing a weak quantum measurement of a solid-state qubit. *Phys. Rev. Lett.* **97,** 166805 (2006).
27. Katz, N. *et al.* Coherent state evolution in a superconducting qubit from partial-collapse measurement. *Science* **312,** 1498–1500 (2006).
28. Katz, N. *et al.* Reversal of the weak measurement of a quantum state in a superconducting phase qubit. *Phys. Rev. Lett.* **101,** 200401 (2008).
29. Kim, Y.-S., Lee, J.-C., Kwon, O. & Kim, Y.-H. Protecting entanglement from decoherence using weak measurement and quantum measurement reversal. *Nature Phys.* **8,** 117–120 (2012).
30. Groen, J. P. *et al.* Partial-measurement back-action and non-classical weak values in a superconducting circuit. *Phys. Rev. Lett.* **111,** 090506 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H. Wiseman, A. N. Korotkov, E. M. Levenson-Falk and N. Roch for discussions, and R. Vijay for contributions to preliminary experiments. This research was supported in part by the Army Research Office, Office of Naval Research and the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Office. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI or the US government.

Author Contributions K.W.M. and S.J.W. performed the experiment, analysed the data and wrote the manuscript. C.M. had the idea for the experiment and edited the manuscript. All work was carried out under the supervision of I.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.W.M. (katerm@berkeley.edu).

doi:10.1038/nature12514

Quasicrystalline structure formation in a classical crystalline thin-film system

Stefan Förster[1], Klaus Meinel[1], René Hammer[1], Martin Trautmann[1] & Wolf Widdra[1,2]

The discovery of quasicrystals[1]—crystalline structures that show order while lacking periodicity—forced a paradigm shift in crystallography. Initially limited to intermetallic systems[1–4], the observation of quasicrystalline structures has recently expanded to include 'soft' quasicrystals in the fields of colloidal and supermolecular chemistry[5–9]. Here we report an aperiodic oxide that grows as a two-dimensional quasicrystal on a periodic single-element substrate. On a Pt(111) substrate with 3-fold symmetry, the perovskite barium titanate $BaTiO_3$ forms a high-temperature interface-driven structure with 12-fold symmetry. The building blocks of this dodecagonal structure assemble with the theoretically predicted Stampfli–Gähler tiling[10,11] having a fundamental length-scale of 0.69 nm. This example of interface-driven formation of ultrathin quasicrystals from a typical periodic perovskite oxide potentially extends the quasicrystal concept to a broader range of materials. In addition, it demonstrates that frustration at the interface between two periodic materials can drive a thin film into an aperiodic quasicrystalline phase, as proposed previously[12]. Such structures might also find use as ultrathin buffer layers for the accommodation of large lattice mismatches in conventional epitaxy[13].

Modern interface engineering in materials science couples the properties of two different materials by means of an atomically sharp interface to new materials and new properties, as for example the combination of ferroelectric and ferromagnetic layers into multiferroic materials. In this respect interface and surface science studies have so far tried to combine intermetallic quasicrystals (QCs) with common materials either by depositing single elemental adlayers on QC surfaces[14] or by thin-film deposition of QC materials on periodic lattices[15]. We report here an alternative: the formation of a two-dimensional QC phase for the growth of an oxide with preferential pseudocubic structure on an elemental metal substrate with 3-fold symmetry. The observation of an aperiodic QC thin-film structure derived from a periodic $BaTiO_3(111)$ layer grown on a Pt(111) single-crystal substrate demonstrates that the concept of a quasicrystal might be expanded to conventional, otherwise periodic, materials when grown heteroepitaxially.

$BaTiO_3$ is one of the most thoroughly investigated perovskite oxide systems and is also widely used in thin-film applications and oxide heterostructures. In its paraelectric phase it crystallizes in a simple cubic lattice with a 4.0 Å lattice constant. As a result of matching lattice conditions it can be grown epitaxially on many other perovskite substrates and also on metal substrates. On Pt(001) and Pt(111) it grows either as a $BaTiO_3(001)$ or as a $BaTiO_3(111)$ periodic thin film, depending on growth conditions and substrate[16–18]. However, as we report here, an aperiodic QC structure can be formed by annealing a $BaTiO_3(111)$ layer several monolayers (14–30 Å) thick on Pt(111). On being annealed to 1,250 K in ultrahigh vacuum (UHV), the film restructures into a QC thin film 2–4 Å thick between a few thicker $BaTiO_3(111)$ islands (see Extended Data Figs 1 and 2). Because the periodically ordered islands cover only a small surface fraction, any diffraction experiment is most sensitive to the extended QC layer between. Figure 1a, b shows room-temperature low-energy electron diffraction (LEED) images for this layer at two different electron energies. Both yield a 12-fold diffraction pattern, which is the signature of a QC layer with dodecagonal symmetry. For comparison a scheme of an ideal 12-fold diffraction pattern is shown in Fig. 1c. To describe a two-dimensional dodecagonal lattice, four lattice vectors are needed; these are represented by the four red arrows in Fig. 1c. These vectors point to the first-order diffraction spots, which are drawn as large solid circles and are also highlighted with red circles in the measured diffraction pattern. In the higher-energy LEED pattern in Fig. 1a these first-order spots are the most

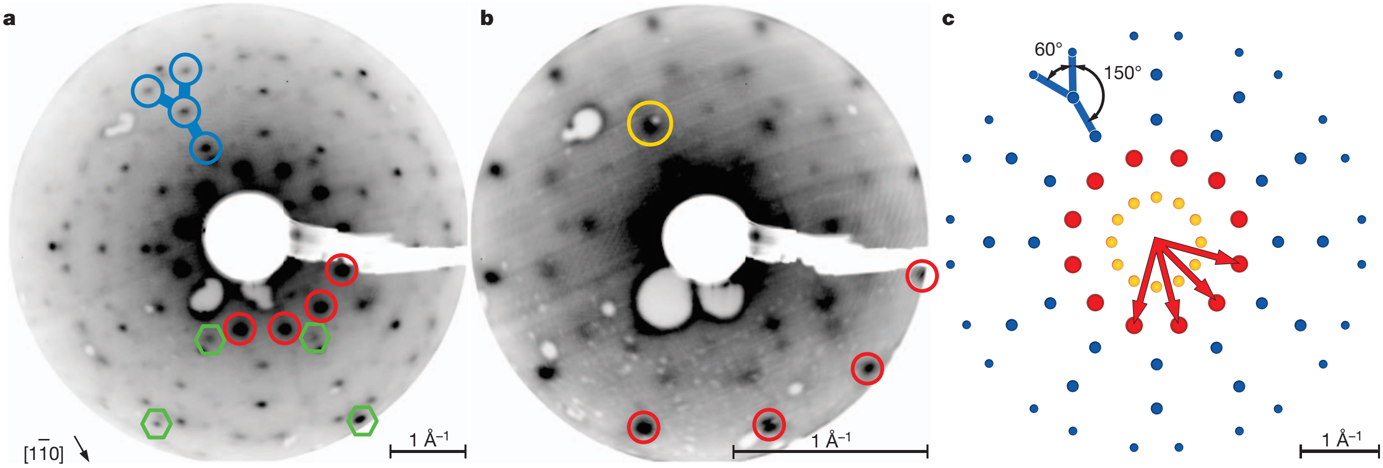


Figure 1 | Electron diffraction from dodecagonal oxide thin film. a, b, LEED patterns of the dodecagonal $BaTiO_3$-derived thin film phase on Pt(111) at electron energies of 66 eV (**a**) and 8 eV (**b**). **c**, Scheme of the 12-fold diffraction pattern, generated by linear combinations of the four lattice vectors, which are indicated by the red arrows. The size of the diffraction spots reflects the diffraction order. Corresponding orders of diffraction are colour-coded identically in all three images. The small green hexagons in **a** mark diffraction from additional (111)-oriented $BaTiO_3$ islands.

[1]Institute of Physics, Martin-Luther-Universität Halle-Wittenberg, 06120 Halle, Germany. [2]Max-Planck-Institut für Mikrostrukturphysik, 06120 Halle, Germany.

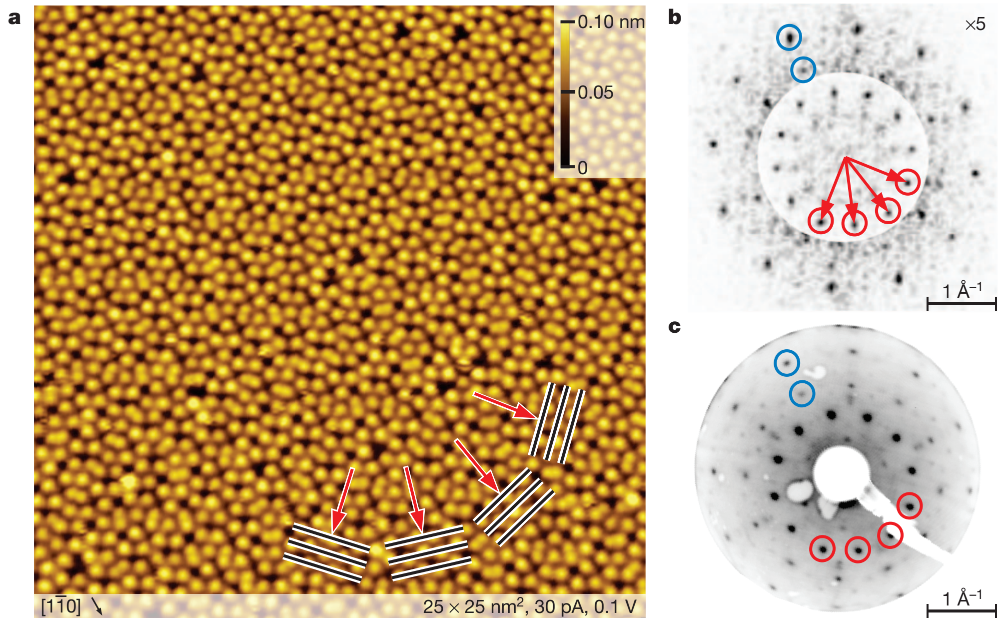


Figure 2 | STM images of 12-fold structure on local scale. **a**, **b**, Low-temperature STM image of the dodecagonal $BaTiO_3$-derived thin film phase on Pt(111) (**a**), and its Fourier transform (**b**). The Bragg planes of four first-order diffraction spots (arrows in **b**) are indicated in **a**. **c**, The corresponding LEED pattern, showing all the characteristic features of the FFT in **b** (electron energy 35 eV). The contrast of the outer part of **b** has been enhanced fivefold.

intense ones. In addition, bright and equidistant spots are visible, arranged in groups of four to form a Y shape with relative angles of 60° and 150° as indicated in Fig. 1a, c. These groups of four spots appear with 12-fold symmetry in the pattern of Fig. 1a. The lower-energy diffraction pattern in Fig. 1b shows, in addition to the very prominent first-order spots (marked by red circles), the inner ring as indicated in orange in Fig. 1b, c. Except for the weak first-order and second-order spots of the $BaTiO_3$(111)-(1 × 1) islands (marked in green in Fig. 1a), the full LEED pattern can be assigned to the dodecagonal structure. However, for a quantitative analysis of the LEED pattern, multiple electron scattering also has to be taken into account, which includes backscattering from the 3-fold substrate (see Extended Data Fig. 5).

An alternative real-space approach by scanning tunnelling microscopy (STM) is presented in Fig. 2, which shows an STM image of the QC layer recorded at 77 K. The STM data reveal a flat film with apparent height variations of 0.1 nm. Within the film the STM resolves an arrangement of protrusions, which are assigned to individual surface atoms. The average distance between the surface atoms is about 6.9 Å. The dominant local structures are triangular or quadratic atomic arrangements with bond angles of roughly 60° and 90°, respectively. Although the film seems ordered, no periodic overlayer structure can be found. In contrast, the fast Fourier transform (FFT) of the STM data (Fig. 2b) contains sharp diffraction spots. The FFT of the STM data reveals three diffraction orders with a clear 12-fold discrete diffraction diagram. Indeed, the FFT resembles exactly the electron diffraction pattern of the QC layer, shown for comparison in Fig. 2c. Both patterns have an inner ring of 12 spots (marked by red circles), which is related to the first-order diffraction spots as discussed above. The real-space Bragg planes responsible for the first-order spots are indicated by black lines in Fig. 2a. The corresponding four reciprocal lattice vectors, which are perpendicular to the Bragg planes, are indicated by the red arrows. Note that the FFT of the STM data from differently selected small areas always produces the full 12-fold diffraction; this excludes the possibility that the observed 12-fold diffraction is the result of a superposition of 4-fold or 6-fold symmetric domains. The outer spots in Fig. 2b (marked blue) are part of the Y-like arrangement of spots in the LEED pattern (blue in Fig. 1). In particular the outermost ring of spots marked in Fig. 2b, c is incompatible with domains of two hexagonal or three quadratic structures that are each rotated by 30°.

To build up a dodecagonal lattice in real space one needs a distinct distance that occurs in the system with the same frequency in all 12 directions. This can ideally be realized by covering the area with equilateral triangles and squares as structural elements[19]. Because triangles and squares enclose angles of 60° and 90°, respectively, the combination of both leads to a distribution of the edges along 12 directions at 30° intervals. The ideal 12-fold arrangement is illustrated in Fig. 3a. The squares and triangles form adjacent dodecagons by having six triangles in the centre, surrounded by an alternating arrangement of six squares and triangles. For the central part this arrangement of triangles and squares is indicated in dark blue and yellow. The area in between adjacent dodecagons is also filled by squares and triangles. The ideal lattice follows a strict rule that leads to a distinct self-similarity. The centre of each dodecagon represents the corner of a similar, but by a factor of $2+\sqrt{3}$ larger dodecagonal structure of squares and triangles, as emphasized by the red lines in Fig. 3a.

We apply this concept to the $BaTiO_3$-derived thin-film structure. Figure 3b, c shows two close-ups with different length scales for the STM image in Fig. 2a. In the high-resolution STM image in Fig. 3c one can clearly identify the local quadratic and triangular arrangement of surface atoms with an average distance of about 6.9 Å between them. Half of the triangles and squares are aligned along the ⟨110⟩ directions, which correspond to the directions of close-packed Pt atoms at the oxide/metal interface. They form the first-generation dodecagons. In contrast with the concept of a 12-fold QC structure exclusively based on quadratic and triangular building blocks as discussed above, the STM images in Fig. 3b, c reveal that an additional third element is involved, namely a rhombus with an inclusion angle of 30°. In fact, this rhombus is a theoretically and experimentally well-established element in QC structures[10,11,19]. As indicated in Fig. 3c, most dodecagons are formed by 12 triangles, 5 squares and 2 rhombi, where the locations of the rhombi seem to vary statistically. Such dodecagons are the building blocks of the well-known Stampfli–Gähler tiling[10,11]. To identify the first-generation dodecagons, their outer ring is highlighted in yellow. These dodecagons with a diameter of 2.7 nm again form triangular and quadratic arrangements on a $2+\sqrt{3}$ larger scale (Fig. 3b). However, even though the positions of the rhombi appear statistically in the interior of these dodecagons, the QC arrangement of the dodecagons is again regular, and higher hierarchies are well developed. On the

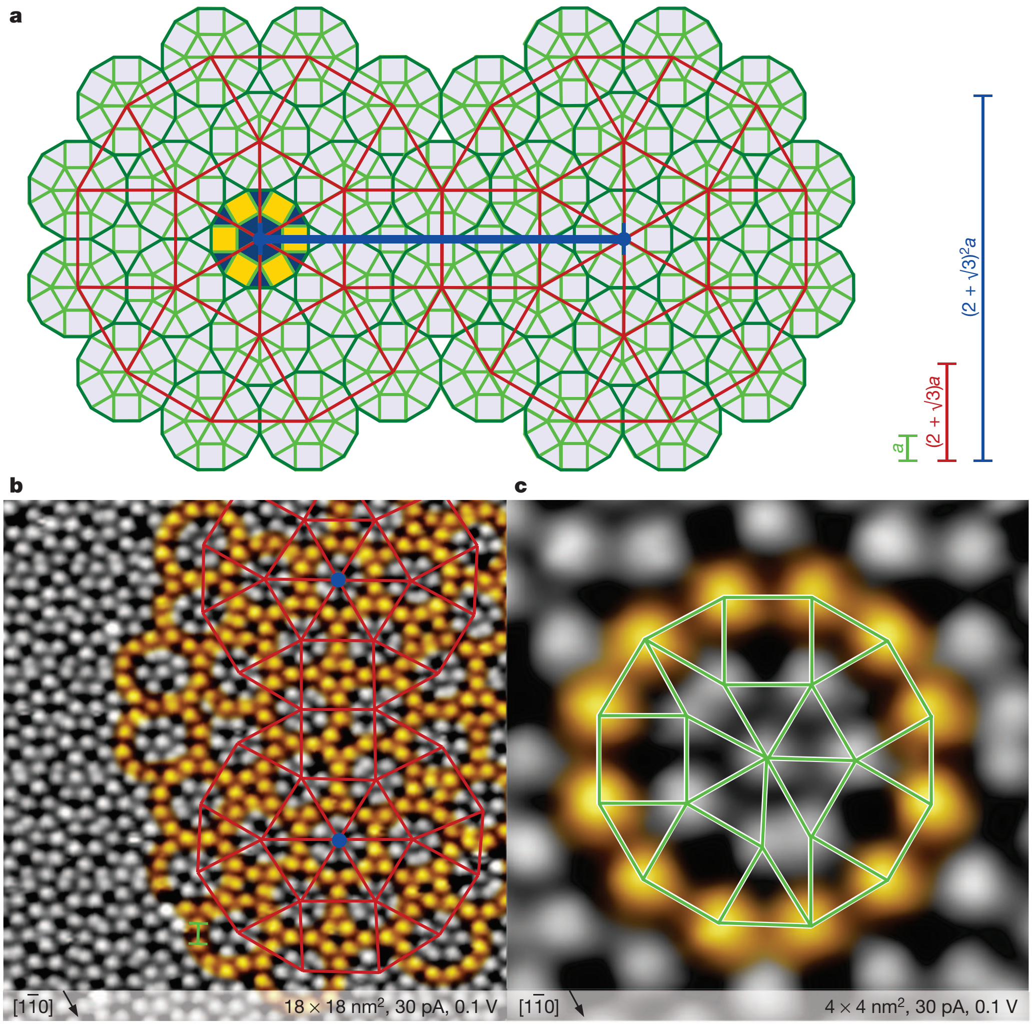


Figure 3 | Dodecagonal tiling as measured by STM. **a**, Ideal dodecagonal tiling built up from equilateral triangles and squares (green lines). Adjacent dodecagons form a self-similar next-generation tiling (red lines) on a $2+\sqrt{3}$ longer length scale, as indicated at the right. The beginning of the third-generation tiling is indicated in blue. **b**, **c**, High-resolution STM images of the $BaTiO_3$-derived dodecagonal QC phase. The second-generation (**b**) and first-generation (**c**) tilings follow the colour coding in **a**. To guide the eye, the outer rings of the first-generation dodecagons are coloured yellow. Inside the dodecagons the atomic arrangement as visible in STM is described by triangles, squares and also rhombi, as marked in **c**.

larger scale again, dodecagons are assembled with a diameter of 10.1 nm. Those dodecagons are forming a still higher hierarchical tiling, now enlarged by a factor of $(2+\sqrt{3})^2$. These higher hierarchical structures can form because there are only few defects in the film. In the 290 nm^2 area shown in Fig. 3b only seven missing atoms can be counted. The STM data show that the experimentally observed QC structure follows, to a large extent, the building concept of an ideal two-dimensional QC. However, depending on the preparation conditions there also exist regions on the surface where the tiling seems more like a random dodecagonal lattice, which means that higher hierarchy tilings are not established.

From the measured LEED data the distance between the surface atoms of the $BaTiO_3$-derived QC layer defining the edge length of the dodecagonal tiling was determined as 6.85 Å, within an accuracy of 0.1 Å (see Extended Data Fig. 4). However, the detailed atomic structure within the triangular, quadratic and rhomboidal building blocks is still unknown. So far the characteristic length of about 6.9 Å and the orientation of the building blocks relative to the substrate are the only hints to a deeper understanding of the observed QC structure. At present one might speculate that the origin of the QC structure is derived from a $BaTiO_3(110)$-c(2×2) layer. In the c(2×2) structure the Ti ions form rhombi with the correct edge length of 6.9 Å. A compression of the rhombus along the shorter diagonal by 14% results in a equilateral triangle, whereas a stretching along the same diagonal by 18% turns the rhombus into a square. Alternating squares and triangles might help to compensate for strain. Clearly, additional theoretical and experimental approaches are needed to solve this unusual structure.

The observation of quasicrystalline structure formation in the heteroepitaxy of otherwise periodic materials paves the way for new quasicrystalline phases. The concept to drive a cubic or quasicubic perovskite into a QC phase on interaction with an appropriate 3-fold substrate might open a wide field of oxidic QCs with many unexpected properties. For that purpose it will be important to continue the growth, to achieve the formation of thicker QC structures.

METHODS SUMMARY

All experiments were performed under UHV conditions. The $BaTiO_3$ films were deposited by radiofrequency-assisted magnetron sputtering on Pt(111) at room temperature. On annealing the sample in oxygen atmosphere, long-range ordered $BaTiO_3(111)$ films develop, as proved by LEED. The QC layer is formed after additional UHV heating at 1,250 K. Equivalent aperiodic thin-film structures were grown by a single annealing step to 1,250 K. Similar aperiodic QC thin-film structures were grown by magnetron sputtering at elevated sample temperatures of 850 K. Alternatively, the QC thin film structure was also grown by molecular beam epitaxy (MBE) on Pt(111) in another UHV system by simultaneous deposition of BaO and Ti in an O_2 atmosphere. For sputtering at elevated temperatures as well as for MBE, a final UHV annealing step to 1,070 K was sufficient to prepare the QC structure. For low-temperature STM the sample was transferred under UHV

conditions to an additional UHV chamber equipped with a home-built STM operated at 77 K.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 July 2012; accepted 30 July 2013.

1. Shechtman, D., Blech, I., Gratias, D. & Cahn, J. Metallic phase with long-range orientational order and no translational symmetry. *Phys. Rev. Lett.* **53,** 1951–1953 (1984).
2. Wang, N., Chen, H. & Kuo, K. H. Two-dimensional quasi-crystal with eightfold rotational symmetry. *Phys. Rev. Lett.* **59,** 1010–1013 (1987).
3. Bendersky, L. Quasicrystal with one-dimensional translational symmetry and a tenfold rotation axis. *Phys. Rev. Lett.* **55,** 1461–1463 (1985).
4. Ishimasa, T., Nissen, H. U. & Fukano, Y. New ordered state between crystalline and amorphous in Ni-Cr particles. *Phys. Rev. Lett.* **55,** 511–513 (1985).
5. Dotera, T. Quasicrystals in soft matter. *Isr. J. Chem.* **51,** 1197–1205 (2011).
6. Zeng, X. B. *et al.* Supramolecular dendritic liquid quasicrystals. *Nature* **428,** 157–160 (2004).
7. Hayashida, K., Dotera, T., Takano, A. & Matsushita, Y. Polymeric quasicrystal: mesoscopic quasicrystalline tiling in ABC star polymers. *Phys. Rev. Lett.* **98,** 195502 (2007).
8. Talapin, D. V. *et al.* Quasicrystalline order in self-assembled binary nanoparticle superlattices. *Nature* **461,** 964–967 (2009).
9. Mikhael, J., Roth, J., Helden, L. & Bechinger, C. Archimedean-like tiling on decagonal quasicrystalline surfaces. *Nature* **454,** 501–504 (2008).
10. Stampfli, P. A dodecagonal quasi-periodic lattice in two dimensions. *Helv. Phys. Acta* **59,** 1260–1263 (1986).
11. Gähler, F. in *Quasicrystalline Materials* (eds Janot, D. & Dubois, J. M.) 272–284 (World Scientific, 1988).
12. Smith, A. R., Chao, K.-J., Niu, Q. & Shih, C.-K. Formation of atomically flat silver films on GaAs with a ‘silver mean’ quasi periodicity. *Science* **273,** 226–228 (1996).
13. Franke, K. J., Gille, P., Rieder, K.-H. & Theis, W. Achieving epitaxy between incommensurate materials by quasicrystalline interlayers. *Phys. Rev. Lett.* **99,** 036103 (2007).
14. Franke, K. J. *et al.* Quasicrystalline epitaxial single element monolayers on icosahedral Al-Pd-Mn and decagonal Al-Ni-Co quasicrystal surfaces. *Phys. Rev. Lett.* **89,** 156104 (2002).
15. Li, G. H., Zhang, D., Jiang, H., Lai, W. & Liu, W. Fully oriented decagonal quasicrystalline thin films on crystalline substrates. *Appl. Phys. Lett.* **71,** 897–899 (1997).
16. Förster, S. & Widdra, W. Growth, structure, and thermal stability of epitaxial $BaTiO_3$ films on Pt(111). *Surf. Sci.* **604,** 2163–2169 (2010).
17. Förster, S., Huth, M., Schindler, K.-M. & Widdra, W. Epitaxial $BaTiO_3$ (100) films on Pt(100): a low-energy electron diffraction, scanning tunneling microscopy, and x-ray photoelectron spectroscopy study. *J. Chem. Phys.* **135,** 104701 (2011).
18. Förster, S., Meinel, K., Schindler, K.-M. & Widdra, W. Oxygen-deficient ordered phases of ultrathin $BaTiO_3$ films on Pt(111). *Surf. Interface Anal.* **44,** 628–634 (2012).
19. Niizeki, N. & Mitani, H. Two-dimensional dodecagonal quasi-lattices. *J. Phys. Math. Gen.* **20,** L405–L410 (1987).

Acknowledgements We thank P. A. Thiel for discussions, R. Kulla for technical support, and E. M. Zollner for assisting in sample preparation. Financial support was provided by Deutsche Forschungsgemeinschaft Sonderforschungsbereich 762 ‘Functionality of Oxidic Interfaces’.

Author Contributions S.F. designed the experiment and performed the sample preparation by radiofrequency magnetron sputtering, room-temperature STM, X-ray photoelectron spectroscopy and LEED measurements. K.M. performed the molecular beam evaporation experiments, Auger electron spectroscopy and spot-profile analysis LEED measurements. R.H. and M.T. contributed the low-temperature STM measurements. S.F., K.M. and W.W. discussed the results and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.W. (wolf.widdra@physik.uni-halle.de).

doi:10.1038/nature12574

Bottom-up control of geomagnetic secular variation by the Earth's inner core

Julien Aubert[1], Christopher C. Finlay[2] & Alexandre Fournier[3]

Temporal changes in the Earth's magnetic field, known as geomagnetic secular variation, occur most prominently at low latitudes in the Atlantic hemisphere[1,2] (that is, from −90 degrees east to 90 degrees east), whereas in the Pacific hemisphere there is comparatively little activity. This is a consequence of the geographical localization of intense, westward drifting, equatorial magnetic flux patches at the core surface[3]. Despite successes in explaining the morphology of the geomagnetic field[4], numerical models of the geodynamo have so far failed to account systematically for this striking pattern of geomagnetic secular variation. Here we show that it can be reproduced provided that two mechanisms relying on the inner core are jointly considered. First, gravitational coupling[5] aligns the inner core with the mantle, forcing the flow of liquid metal in the outer core into a giant, westward drifting, sheet-like gyre[6]. The resulting shear concentrates azimuthal magnetic flux at low latitudes close to the core–mantle boundary, where it is expelled by core convection and subsequently transported westward. Second, differential inner-core growth[7,8], fastest below Indonesia[6,9], causes an asymmetric buoyancy release in the outer core which in turn distorts the gyre, forcing it to become eccentric, in agreement with recent core flow inversions[6,10,11]. This bottom-up heterogeneous driving of core convection dominates top-down driving from mantle thermal heterogeneities, and localizes magnetic variations in a longitudinal sector centred beneath the Atlantic, where the eccentric gyre reaches the core surface. To match the observed pattern of geomagnetic secular variation, the solid material forming the inner core must now be in a state of differential growth rather than one of growth and melting induced by convective translation[7,8].

The peculiar geographical localization (both in latitude and longitude, Fig. 1a and b, respectively) of geomagnetic secular variation observed during the historical era[1] (AD 1590–1990) has now been mapped to an unprecedented level of accuracy by virtue of more than one decade of global and continuous satellite magnetic observations. Recent models[12–14] benefit from improved removal of the contribution from external fields, and are now reliably able to resolve structures down to a 1,500-km lateral extent at the Earth's core–mantle boundary (spherical harmonic degree 13, Fig. 2a, b). To exploit this wealth of data fully, and uncover the dynamical processes operating in the Earth's core, it is necessary to combine magnetic observations with numerical models of the geodynamo[15]. These have steadily improved over the past two decades, with the standard models now well understood, thanks to concerted benchmarking activities and systematic exploration of the accessible parameter space[16]. Despite this progress, success in understanding the geomagnetic secular variation and predicting its future evolution remains conditional on our ability to explain its detailed structure with physical mechanisms simple and robust enough to withstand extrapolation from the simplified working regime of current models to the conditions of Earth's core. An outstanding difficulty for standard models has been how to reproduce the westward drift of low-latitude magnetic flux patches at the core–mantle boundary[3,14] in a self-consistent fashion. Quasi-steady westward drift has previously been obtained[17–19], but it becomes erratic[16] in models where magnetic field advection becomes realistically strong relative to diffusion. Furthermore, the Atlantic–Pacific asymmetry in the kinematics of these patches suggests some form of heterogeneous boundary control[6]. The combined effect on secular variation of top-down forcing from lower-mantle thermal anomalies[20] versus bottom-up forcing from the recently discovered inner-core translational instability[7,8] has remained unexplored.

Our coupled Earth numerical dynamo model (Methods) addresses these issues by building on the well understood basis of standard geodynamo models[16] and including some simple ingredients from early

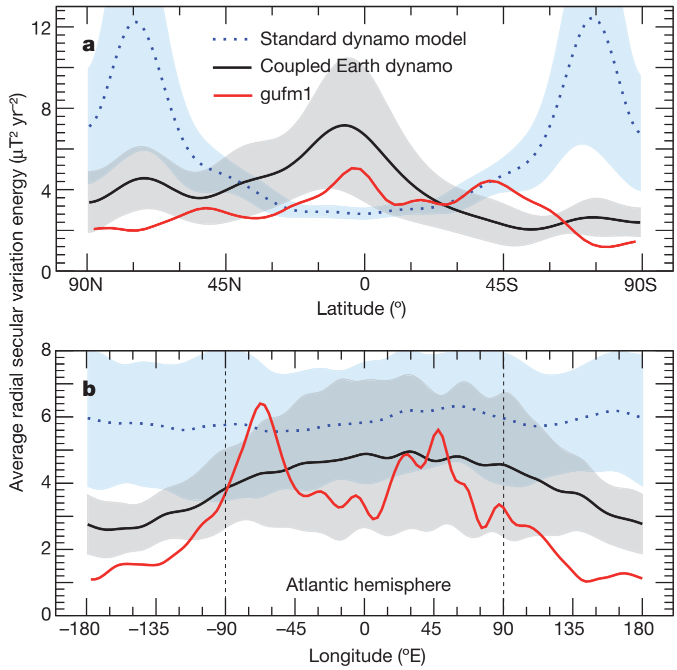


Figure 1 | Geographical localization of geomagnetic secular variation. **a**, **b**, Profiles of the time-averaged radial secular variation energy, averaged over longitude (**a**) or latitude (**b**). In each panel, data are taken from the historical geomagnetic field model[1] gufm1 (red line; spanning AD 1590–1990 and evaluated at the core–mantle boundary), from 400-year time-averages successively taken within a 83,000-year sequence of a standard numerical dynamo model (dotted blue line, average profile; blue shading, ±1 s.d. in the distribution of 400-year time averages), and within a 67,000-year sequence of our coupled Earth dynamo model (black line, average profile; grey shading, ±1 s.d. in the distribution of 400-year time averages). The numerical dynamo outputs are filtered at spherical harmonic degree and order 8, and the gufm1 model is presented at native resolution. The standard model has rigid, electrically insulating and non-moving boundaries. The coupled Earth dynamo model includes visco-magnetic outer-core/inner-core coupling, gravitational inner-core/mantle coupling, and thermochemical, heterogeneous coupling between the outer core, inner core and mantle.

[1]Dynamique des Fluides Géologiques, Institut de Physique du Globe de Paris, Sorbonne Paris Cité, Université Paris Diderot, UMR 7154 CNRS, F-75005 Paris, France. [2]Division of Geomagnetism, National Space Institute, Technical University of Denmark, Elektrovej, 2800 Kongens Lyngby, Denmark. [3]Géomagnétisme, Institut de Physique du Globe de Paris, Sorbonne Paris Cité, Université Paris Diderot, UMR 7154 CNRS, F-75005 Paris, France.

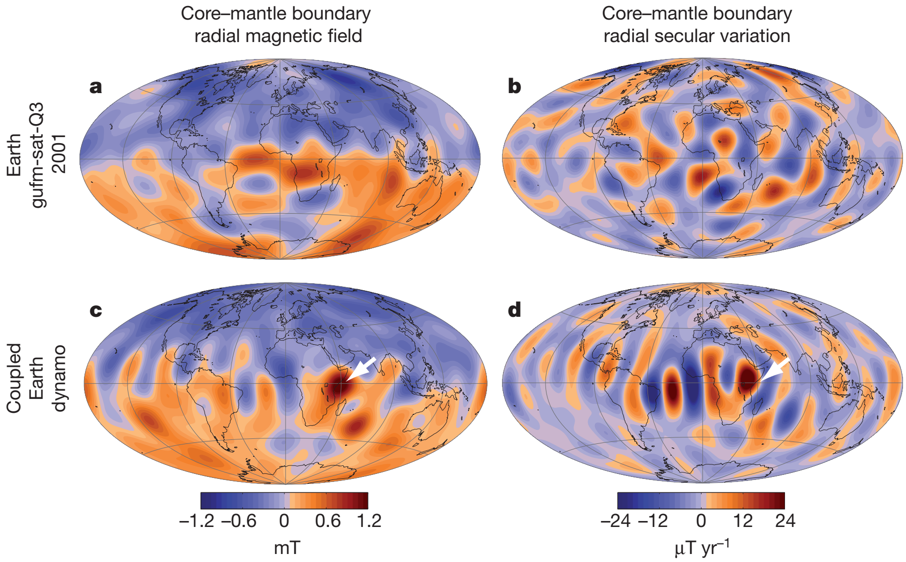


Figure 2 | Maps of the magnetic field and secular variation. **a**–**d**, Atlantic-centred Hammer projections of the core–mantle boundary radial magnetic field (**a**, **c**) and its temporal rate of change or radial secular variation (**b**, **d**); data are from geomagnetic field model[14] gufm-sat-Q3 in 2001 (**a**, **b**) and a snapshot of the coupled Earth dynamo model (**c**, **d**; orange denotes an outward magnetic field), both filtered at spherical harmonic degree and order 13. White arrows in **c** and **d** mark a location chosen for further analysis in Figs 3 and 4.

modelling efforts[17,21,22] that have recently been neglected. A field morphology similar to that of the Earth[4] (Fig. 2 and Extended Data Fig. 1) is achieved by choosing ratios of the timescales for magnetic induction, magnetic diffusion and the length of the day as close as computationally feasible to their Earth counterparts. We compromise on the ratio of the length of the day to the viscous diffusion timescale (the Ekman number, Ek), which, although low, lies an order of magnitude above state-of-the art values[18,19], in order to facilitate a parameter space exploration (Supplementary Information and Extended Data Table 1). The influence of viscosity is further mitigated by adopting a stress-free mechanical condition at the core–mantle boundary.

Robust westward drift and time-dependence of low-latitude secular variation similar to that observed over the past 400 years are achieved in the coupled Earth dynamo model through a mechanism based on indirect exchanges of angular momentum between the outer core and the mantle via the inner core. The inner core couples with the outer core primarily through a magnetic torque, which dominates at the Ekman number studied. At the base of the outer core, thermochemical ‘winds’ resulting from the interaction of the Coriolis force with convective upwellings entrain the inner core in the eastward direction[21,23] with respect to the planetary rotating frame. But a gravitational torque[5,22–24], resulting from coupling between lower-mantle mass anomalies and the induced topography at the inner-core surface, transfers angular momentum between the inner core and mantle. In a situation where it only experiences this gravitational torque, the mantle is then also entrained in the eastward direction. Conservation of angular momentum in the planetary rotating frame then requires a sustained westward flow (see below) with respect to the mantle in the uppermost part of the outer core[6,22]. Both the gravitational torque and the net torque exerted by the outer core on the inner core vanish when averaged over time (Methods). This leads to long-term co-rotation between the mantle, inner core and lower outer core, with thermochemical wind gradients maintaining the westward drift in the upper outer core. For a given convective forcing, this drift is maximized if direct coupling between the outer core and the mantle is much weaker than indirect coupling via the inner core, as is the case in the coupled Earth dynamo model.

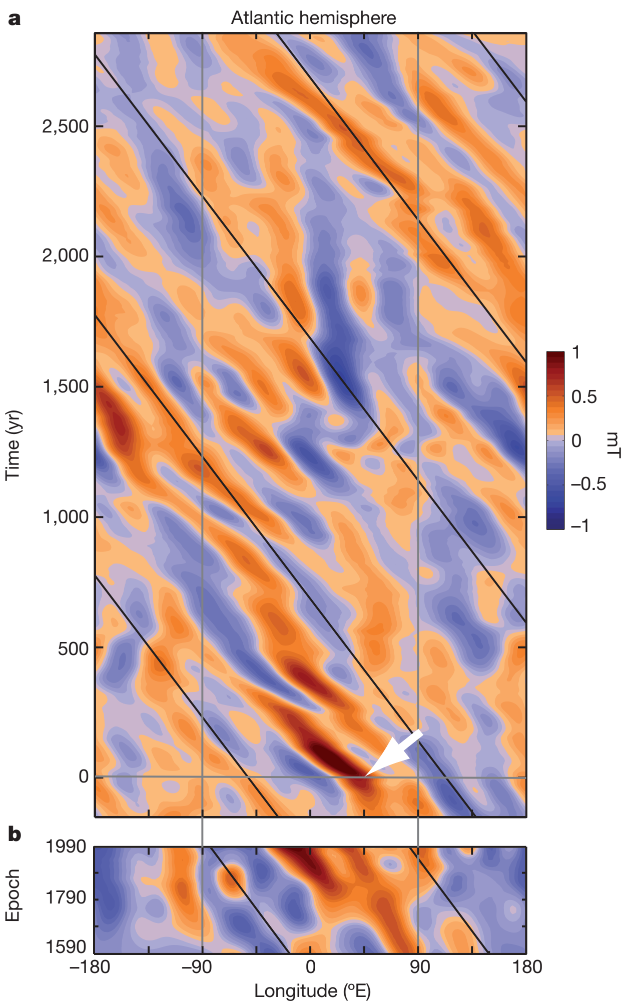


Figure 3 | Temporal evolution of magnetic structures at the Equator. **a**, **b**, Time–longitude plot of the radial magnetic field at the Equator evaluated at the core–mantle boundary (orange denotes an outward magnetic field) from a 3,000-year sequence of the coupled Earth dynamo model filtered to spherical harmonic degree and order 8 (**a**), and from the 400-year sequence of geomagnetic field model[1] gufm1 at its native resolution (**b**). Grey vertical lines mark the Atlantic hemisphere. The grey horizontal line (model time 0) marks the time of the snapshot presented in Figs 2 and 4, and the white arrow marks the reference location. The model time 0 is chosen to be the start of a magnetic patch emergence and drift sequence representative of the localization of secular variation diagnosed in Fig. 1. See Supplementary Video for the evolution of magnetic features at all latitudes in the coupled Earth dynamo model over the entire temporal sequence. The slanted black lines denote a reference westward drift velocity of 10 km yr^{-1}.

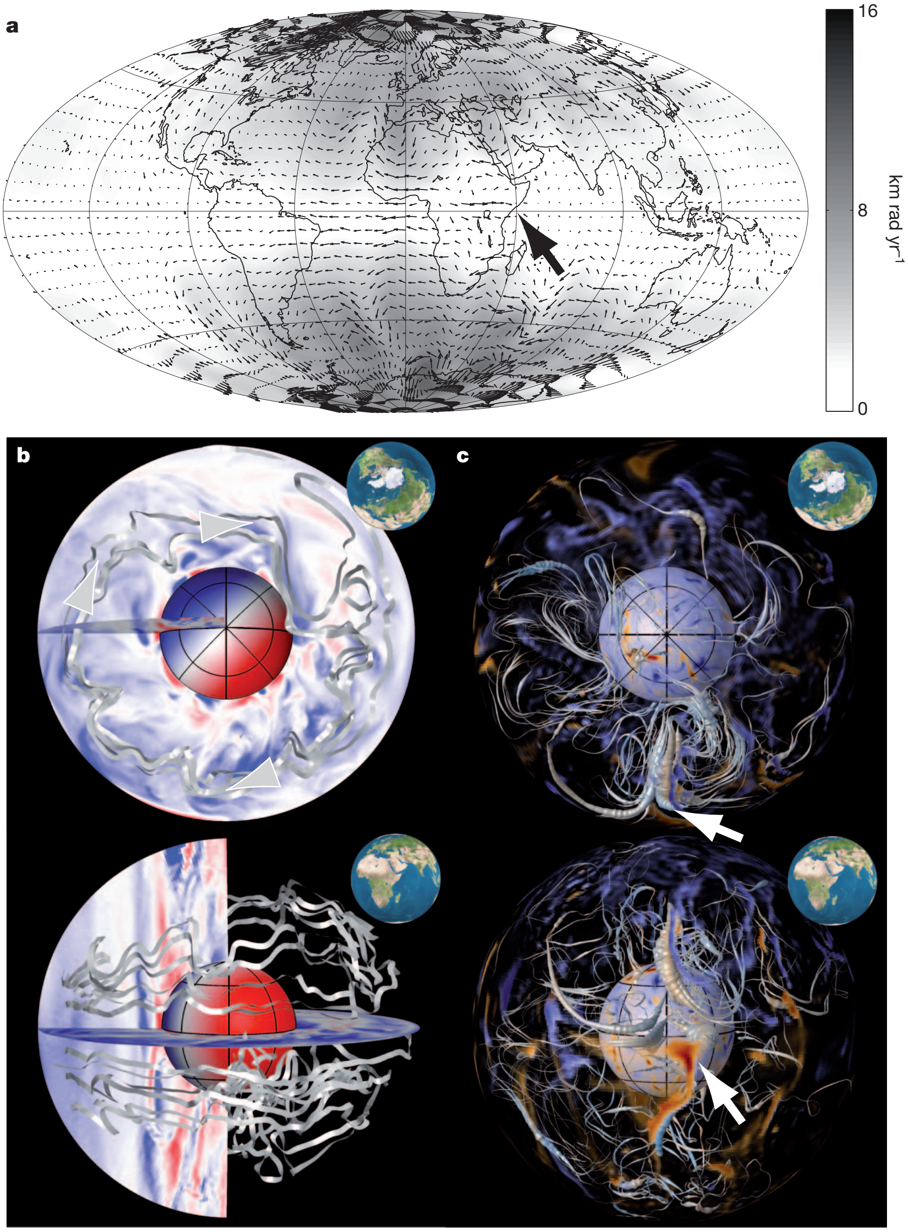


Figure 4 | Internal fluid flow and magnetic structure. **a**, Atlantic-centred Hammer projection of flow at the core–mantle boundary in a snapshot of the coupled Earth dynamo model (small arrows, arbitrary scaling), superimposed on a grey-scale map of the amount of local surface rotation (measured by the absolute toroidal scalar[6], in km rad yr^{-1}). The large black arrow marks the reference location introduced in Figs 2 and 3. **b**, Stream ribbon representations (top, north polar view; bottom, near-equatorial view; insets, indication of the viewing angle) of the internal velocity field structure, superimposed on equatorial and meridional cuts of the azimuthal velocity field (red is eastwards, blue is westwards, maximum 28 km yr^{-1}). Grey arrow heads mark the general flow circulation and the twist of the ribbons is proportional to the local vorticity. The inner-core boundary is colour-coded with the amplitude of the hemispherical modulation in mass anomaly release (red corresponds to excess buoyancy). **c**, Magnetic field line (grey) representations of the internal magnetic structure (same viewing angles as **b**). The field line thickness is proportional to the local magnetic energy density in the shell. Inner and outer boundaries are colour-coded with the amplitude of the unfiltered radial magnetic field (orange is outward, outer boundary is selectively transparent, maxima are 4 mT at the core–mantle boundary and 7 mT at the inner-core boundary). The model snapshot is taken at the same time as in Fig. 2 (which is time 0 in Fig. 3), with white arrows marking the reference location introduced in these figures.

Heterogeneous thermochemical boundary coupling is also included in the coupled Earth dynamo model. We adopt fixed-flux-type[18] conditions for the thermochemical buoyancy both at the core–mantle boundary (where lower-mantle thermal anomalies significantly affect the geodynamo[9,20,25]) and at the inner-core boundary (where solid translation of the inner core[7,8] can spatially modulate the inner-core growth rate, with equally important consequences[6,26,27]). At the core–mantle boundary, a mass anomaly flux pattern[9] derived from lower-mantle seismic tomography[28] is superimposed on a vanishing homogeneous flux. This represents a situation where the upper outer core is neutrally stratified (although our results are robust against a possible stable stratification[29,30], Supplementary Information). At the inner-core boundary, we prescribe a longitudinal hemispheric modulation of mass anomaly release related to inner-core translation. The maximum outward flux is located beneath Indonesia, in resonance with the heterogeneous mantle forcing[9], and as suggested by earlier investigations[6]. Inner-core translation thus proceeds from east to west, opposite to inferences[7] based on upper inner-core seismic texturing. The latter interpretation is however not applicable here according to a parameter space exploration of our model (Supplementary Information and Extended Data Figs 2, 3), which favours

an inner-core translation slower than its mean growth rate. The upper inner-core seismic properties may thus be dominated by solidification texturing[9] rather than texturing induced by the slow ageing of translating iron grains.

By including these elements, the coupled Earth dynamo model is able to produce magnetic variations that are dominated by the emergence of intense, westward-drifting, equatorial flux patches of normal (the same as the ambient) polarity under the Atlantic hemisphere (Supplementary Video). The temporal sequence of emergence, drift and diffusion of these patches agrees between the model and Earth by virtue of the realistic ratio between magnetic induction and diffusion timescales. The associated distribution of average secular variation energy (Fig. 1) quantitatively reproduces the longitudinal and latitudinal localizations of geomagnetic secular variation which standard geodynamo models fail to manage. We find (Extended Data Fig. 4) that latitudinal localization results primarily from the introduction of inner-core/mantle gravitational coupling, whereas longitudinal localization is mainly a consequence of heterogeneous thermochemical coupling between the inner and outer core. Control from the heterogeneous mantle has a secondary effect (Supplementary Information), operating constructively with the inner-core control to promote low secular variation in the Pacific[20]. The instantaneous field and secular variation at the core–mantle boundary exhibit detailed (Fig. 2c, d) and quantitative (Extended Data Fig. 1) agreement with present-day geomagnetic maps derived from high-resolution satellite data (Fig. 2a, b). A time–longitude plot of the evolution of radial magnetic field at the Equator from the coupled Earth dynamo model (Fig. 3a) closely resembles that obtained during the past 400 years (ref. 1; Fig. 3b), with intense patches coherently moving westwards through much of the Atlantic hemisphere, while in the Pacific hemisphere significantly weaker patches undergo slower westward drift with occasional standing and eastward-drifting phases (see also Supplementary Video). The peak of azimuthally moving power (Extended Data Fig. 5) is reached at the Equator, with a coherent westward drift at 14 km yr^{-1}, comparable to the observed speeds[3] of 17 km yr^{-1}. In contrast, the standard dynamo model of Fig. 1 yields a weak and erratic drift at low latitudes with a small amount of power moving slowly eastwards.

The longitudinal localization of magnetic variations in the coupled Earth dynamo model is underlain by a fluid flow (Fig. 4a, b) that agrees remarkably well with recent observation-based inferences of the core flow[6,10,11]. This enables a new, dynamically consistent, interpretation of the physical processes underlying such kinematic flow maps. The gyre driving the westward drift has a sheet-like structure with invariance along the Earth's rotation axis. It produces significant shear at low latitudes close to the core–mantle boundary, concentrating energetic azimuthal magnetic field lines in this region (Fig. 4c). The outer-core upwelling pattern caused by the inner-core hemispherical driving distorts the gyre into an eccentric path that reaches the core–mantle boundary beneath the Atlantic. Expulsion of azimuthal magnetic flux by columnar convection is thus strongest at this location and forms intense radial magnetic flux patches which are subsequently swept westwards by the background flow. The gyre itself is advected westwards, quasi-periodically entering into resonance with the buoyancy heterogeneity of the inner-core boundary. Its shape thus modulates between more eccentric (for example, time 0 in Fig. 3, and Figs 2 and 4) and less eccentric phases with weaker westward drift and magnetic patches more evenly distributed in longitude (for example, time 1,400 in Fig. 3). Assuming that the geodynamo is currently in a phase of more intense gyre eccentricity, our results thus indicate that advection by a planetary-scale thermochemical wind is the origin of low-latitude geomagnetic secular variation. The mechanisms identified in the coupled Earth dynamo model are generic processes related to boundary coupling and involve a competition between the Coriolis, buoyancy and magnetic forces. As realistic relative proportions of these forces are maintained in the simulation over the centennial timescale range relevant for the convective processes (Supplementary Information), these mechanisms should continue to operate in the regime of the Earth's core.

Immediate implications of our findings include the possibility of formulating a data assimilation framework based on a numerical dynamo that includes such boundary couplings. A dynamically consistent prediction of the future geomagnetic field evolution is in sight, and improved knowledge of the past of the core may be accessible through the reanalysis of historical and archaeomagnetic field data[15]. Our results also invite a re-evaluation of the possible strength of inner-core translation with respect to its homogeneous growth. Strong heterogeneous forcing from the inner core is in contradiction with observed geomagnetic secular variation, as it leads to core surface flows and magnetic structures with the reverse drift direction (Extended Data Fig. 3).

METHODS SUMMARY

We solve[31] for Boussinesq convection, thermochemical buoyancy transport and magnetic induction in a system with constant axial rotation at a rate Ω (defining the planetary rotating frame), comprising a spherical fluid shell between radii r_i and r_o ($r_i/r_o = 0.35$), a solid inner core of radius r_i with the same electrical conductivity as the fluid shell, and an insulating solid shell representing the mantle between radii r_o and $1.83r_o$. The inner-core boundary is of a no-slip, electrically conducting type, and axially rotates at a rate Ω_i under the combined influence of viscous, gravitational and magnetic torques[23]. The core–mantle boundary is of a free-slip, electrically insulating type and axially rotates at a rate Ω_m under the influence of the gravitational torque. The gravitational torque on the inner core due to the mantle is given by[22,23] $\Gamma_G = -\Gamma\tau(\Omega_i - \Omega_m)$, where $\Gamma\tau = 2.9 \times 10^{20}$ N m yr, Γ being the gravitational coupling constant and τ the viscous relaxation time of the inner core. The inner-core boundary mass anomaly flux is fixed in the frame rotating with the mantle, with a homogeneous contribution f and a longitudinally hemispherical heterogeneity $\Delta f = 0.8f$ maximal[6] at longitude 90° E. The core–mantle boundary mass anomaly flux is fixed in the frame rotating with the mantle, with a vanishing homogeneous contribution and an heterogeneous pattern[9] derived from lower-mantle seismic tomography[28], of peak-to-peak amplitude $\Delta f_o = 0.115f$. The Ekman number is set to $\mathrm{Ek} = \nu/\Omega D^2 = 3 \times 10^{-5}$ (here ν is the fluid viscosity and $D = r_o - r_i$ is the fluid shell depth). The mass anomaly flux Rayleigh number is $\mathrm{Ra_F} = g_o f/\rho\Omega^3 D^2 = 9.3 \times 10^{-5}$ (here g_o is the gravity at the core–mantle boundary and ρ is the fluid density). The ratios between the fluid viscosity, thermal diffusivity κ and magnetic diffusivity λ are set to $\mathrm{Pr} = \nu/\kappa = 1$ and $\mathrm{Pr_m} = \nu/\lambda = 2.5$ (here Pr and $\mathrm{Pr_m}$ are respectively the Prandtl number and the magnetic Prandtl number). The non-dimensional model output is rescaled to the dimensional world using scaling principles thought to hold in both model and Earth: secular variation time scaling[32] and convective-power-based magnetic field scaling[33].

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 March; accepted 14 August 2013.

1. Jackson, A., Jonkers, A. R. T. & Walker, M. R. Four centuries of geomagnetic secular variation from historical records. *Phil. Trans. R. Soc. Lond. A* **358,** 957–990 (2000).
2. Holme, R., Olsen, N. & Bairstow, F. L. Mapping geomagnetic secular variation at the core-mantle boundary. *Geophys. J. Int.* **186,** 521–528 (2011).
3. Finlay, C. C. & Jackson, A. Equatorially dominated magnetic field change at the surface of Earth's core. *Science* **300,** 2084–2086 (2003).
4. Christensen, U. R., Aubert, J. & Hulot, G. Conditions for Earth-like geodynamo models. *Earth Planet. Sci. Lett.* **296,** 487–496 (2010).
5. Buffett, B. Gravitational oscillations in the length of day. *Geophys. Res. Lett.* **23,** 2279–2282 (1996).
6. Aubert, J. Flow throughout the Earth's core inverted from geomagnetic observations and numerical dynamo models. *Geophys. J. Int.* **192,** 537–556 (2013).
7. Monnereau, M., Calvet, M., Margerin, L. & Souriau, A. Lopsided growth of Earth's inner core. *Science* **328,** 1014–1017 (2010).
8. Alboussière, T., Deguen, R. & Melzani, M. Melting-induced stratification above the Earth's inner core due to convective translation. *Nature* **466,** 744–747 (2010).
9. Aubert, J., Amit, H., Hulot, G. & Olson, P. Thermochemical flows couple the Earth's inner core growth to mantle heterogeneity. *Nature* **454,** 758–761 (2008).
10. Pais, M. A. & Jault, D. Quasi-geostrophic flows responsible for the secular variation of the Earth's magnetic field. *Geophys. J. Int.* **173,** 421–443 (2008).
11. Gillet, N., Pais, M. A. & Jault, D. Ensemble inversion of time-dependent core flow models. *Geochem. Geophys. Geosyst.* **10,** Q06004 (2009).
12. Olsen, N., Mandea, M., Sabaka, T. J. & Toffner-Clausen, L. The CHAOS-3 geomagnetic field model and candidates for the 11th generation IGRF. *Earth Planets Space* **62,** 719–727 (2010).
13. Lesur, V., Wardinski, I., Hamoudi, M. & Rother, M. The second generation of the GFZ Reference Internal Magnetic Model: GRIMM-2. *Earth Planets Space* **62,** 765–773 (2010).

14. Finlay, C. C., Jackson, A., Gillet, N. & Olsen, N. Core surface magnetic field evolution 2000–2010. *Geophys. J. Int.* **189,** 761–781 (2012).
15. Fournier, A. *et al.* An introduction to data assimilation and predictability in geomagnetism. *Space Sci. Rev.* **155,** 247–291 (2010).
16. Christensen, U. R. & Wicht, J. in *Treatise on Geophysics* Vol. 8, *Core Dynamics* (ed. Schubert, G.) Ch. 8 (Elsevier, 2007).
17. Kuang, W. & Bloxham, J. An Earth-like numerical dynamo model. *Nature* **389,** 371–374 (1997).
18. Sakuraba, A. & Roberts, P. H. Generation of a strong magnetic field using uniform heat flux at the surface of the core. *Nature Geosci.* **2,** 802–805 (2009).
19. Miyagoshi, T., Kageyama, A. & Sato, T. Zonal flow formation in the Earth's core. *Nature* **463,** 793–796 (2010).
20. Christensen, U. R. & Olson, P. Secular variation in numerical geodynamo models with lateral variations of boundary heat flux. *Phys. Earth Planet. Inter.* **138,** 39–54 (2003).
21. Glatzmaier, G. A. & Roberts, P. H. Rotation and magnetism of Earth's inner core. *Science* **274,** 1887–1891 (1996).
22. Buffett, B. A. & Glatzmaier, G. A. Gravitational braking of inner-core rotation in geodynamo simulations. *Geophys. Res. Lett.* **27,** 3125–3128 (2000).
23. Aubert, J. & Dumberry, M. Steady and fluctuating inner core rotation in numerical geodynamo models. *Geophys. J. Int.* **184,** 162–170 (2011).
24. Dumberry, M. Geodynamic constraints on the steady and time-dependent inner core axial rotation. *Geophys. J. Int.* **170,** 886–895 (2007).
25. Gubbins, D., Sreenivasan, B., Mound, J. & Rost, S. Melting of the Earth's inner core. *Nature* **473,** 361–363 (2011).
26. Davies, C. J., Silva, L. & Mound, J. E. On the influence of a translating inner core in models of outer core convection. *Phys. Earth Planet. Inter.* **214,** 104–114 (2013).
27. Olson, P. & Deguen, R. Eccentricity of the geomagnetic dipole caused by lopsided inner core growth. *Nature Geosci.* **5,** 565–569 (2012).
28. Masters, G., Laske, G., Bolton, H. & Dziewonski, A. in *Earth's Deep Interior* (eds Karato, S., Forte, A., Liebermann, R. C., Masters, G. & Stixrude, L.) 63–87 (AGU Monograph Vol. 117, American Geophysical Union, 2000).
29. Helffrich, G. & Kaneshima, S. Outer-core compositional stratification from observed core wave speed profiles. *Nature* **468,** 807–810 (2010).
30. Gubbins, D. & Davies, C. J. The stratified layer at the core-mantle boundary caused by baro-diffusion of oxygen, sulphur and silicon. *Phys. Earth Planet. Inter.* **215,** 21–28 (2013).
31. Aubert, J., Labrosse, S. & Poitou, C. Modelling the palaeo-evolution of the geodynamo. *Geophys. J. Int.* **179,** 1414–1428 (2009).
32. Lhuillier, F., Fournier, A., Hulot, G. & Aubert, J. The geomagnetic secular-variation timescale in observations and numerical dynamo models. *Geophys. Res. Lett.* **38,** L09306 (2011).
33. Christensen, U. R. & Aubert, J. Scaling properties of convection-driven dynamos in rotating spherical shells and application to planetary magnetic fields. *Geophys. J. Int.* **166,** 97–114 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the French Agence Nationale de la Recherche (grant ANR-2011-BS56-011). Numerical computations were performed at S-CAPAD, IPGP, France, and using HPC resources from GENCI-IDRIS (grants 2012-042122 and at 2013-042122). This is IPGP contribution 3419.

Author Contributions J.A. designed the project and carried out the numerical experiments. J.A. and C.C.F. designed the numerical experiments and processed the results. J.A, C.C.F. and A.F. discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A. (aubert@ipgp.fr).

doi:10.1038/nature12525

Key role of symbiotic dinitrogen fixation in tropical forest secondary succession

Sarah A. Batterman[1], Lars O. Hedin[1], Michiel van Breugel[2], Johannes Ransijn[3,4], Dylan J. Craven[5]† & Jefferson S. Hall[2]

Forests contribute a significant portion of the land carbon sink, but their ability to sequester CO_2 may be constrained by nitrogen[1–6], a major plant-limiting nutrient. Many tropical forests possess tree species capable of fixing atmospheric dinitrogen (N_2)[7], but it is unclear whether this functional group can supply the nitrogen needed as forests recover from disturbance or previous land use[1], or expand in response to rising CO_2 (refs 6, 8). Here we identify a powerful feedback mechanism in which N_2 fixation can overcome ecosystem-scale deficiencies in nitrogen that emerge during periods of rapid biomass accumulation in tropical forests. Over a 300-year chronosequence in Panama, N_2-fixing tree species accumulated carbon up to nine times faster per individual than their non-fixing neighbours (greatest difference in youngest forests), and showed species-specific differences in the amount and timing of fixation. As a result of fast growth and high fixation, fixers provided a large fraction of the nitrogen needed to support net forest growth (50,000 kg carbon per hectare) in the first 12 years. A key element of ecosystem functional diversity was ensured by the presence of different N_2-fixing tree species across the entire forest age sequence. These findings show that symbiotic N_2 fixation can have a central role in nitrogen cycling during tropical forest stand development, with potentially important implications for the ability of tropical forests to sequester CO_2.

The recognition that large proportions of tropical forests are recovering from cutting, agricultural land use or natural disturbances[1,9,10] raises fundamental questions about limits on the tropical carbon sink. Nitrogen may constrain the post-disturbance recovery of vegetation and soil carbon pools[1,3,11,12] and may limit forest response[6,8] to rising CO_2, but the extent to which biological N_2 fixation can relieve these constraints remains unresolved[3,13,14]. Tree species capable of symbiotic N_2 fixation are abundant in many tropical forests, but recent observations show that, depending on local conditions, individual trees may or may not produce the symbiotic nodules where fixation occurs[15], that phosphorus may ultimately constrain fixation in tropical soils[8,16–18] and that some tropical forests contain a low diversity of N_2-fixing species[7,19].

It is therefore crucial to resolve whether N_2-fixing trees can relieve nitrogen limitation and ensure fixation as an ecosystem service[20] during tropical forest secondary succession. The existence of a feedback between post-disturbance forest regrowth and N_2 fixation could influence not only the limits to forest CO_2 uptake but also how we understand and model potentially nonlinear interactions between land use, nitrogen cycling and the land carbon sink[2,3,21].

Here we use a replicated, natural ecosystem experiment of forest recovery from land use across Panamanian forests to ask, first, whether symbiotic N_2 fixation can overcome the nitrogen deficiency imposed by rapid biomass accumulation during secondary succession; and, second, whether ecosystem-scale fixation depends on the identity and diversity of N_2-fixing species.

The Agua Salud Project encompasses approximately 15 km^2 of tropical moist forests[22] on phosphorus-poor oxisols similar to soils in large tropical regions worldwide[23] (Supplementary Note 1 and Supplementary Table 1). We examined 12 randomly chosen 0.2-ha plots abandoned from pasture 5, 12 or 30 years ago (four replicate plots per age), plots abandoned approximately 80 years ago (two 1.0-ha plots), and forests near mature status (approximately 300 years; two 1.0-ha plots).

We measured diameters of 13,372 trees from 297 identified tree species, and collected 244 trees to develop allometric biomass equations for the 26 dominant species[22] (Methods, Supplementary Note 2 and Supplementary Table 2). Fixers constituted 2.6–14.5% of basal area (Table 1), which is at the low end for many tropical forests, and several of our most abundant species are widespread throughout the neotropics[7]. Comparison between 2009 and 2011 measures enabled us to quantify directly individual tree growth rates and changes in community composition. Biomass carbon pools were calculated from tree allometries (sum of leaves, wood and roots; Supplementary Note 2, Supplementary Fig. 1 and Supplementary Table 2); leaf nutrients across 55 species were similar to other forests[24].

We determined fixation in 268 trees (approximately 30% of fixers in all plots) across the nine most abundant species (>70% of fixer basal area across plots; Supplementary Notes 3 and 4). For each tree we quantified active N_2-fixing nodules beneath the canopy (nodulation; $g\,m^{-2}$) and estimated the fixation rate based on ^{15}N-tracer incubations or acetylene reduction (Supplementary Note 3 and Supplementary Table 4). This enabled us to scale up fixation to the ecosystem, by summing across all individual N_2 fixers in all plots using species- and forest-age-specific data on nodulation and per-nodule biomass fixation (scenario number 4 in Supplementary Note 3).

We observed rapid forest regrowth and biomass carbon accumulation across the chronosequence (Fig. 1a and Supplementary Table 3). The rate of biomass accumulation (that is, net biomass carbon increase per area and forest age) was exceptionally high in young forests, such that our 12-year forests contained >40% (>50,000 $kg\,C\,ha^{-1}$; >4,100 $kg\,C\,ha^{-1}\,yr^{-1}$) of the biomass present in the 300-year forests (approximately 120,500 $kg\,C\,ha^{-1}$). Biomass carbon accumulation decelerated in older forests (black line, Fig. 1a), with a theoretical long-term equilibrium biomass that compared well with eleven old-growth Panamanian forests[25] (red line, Fig. 1a).

We also observed a dramatic trend in N_2 fixation across the forest chronosequence (Fig. 1b, Supplementary Fig. 2 and Supplementary Table 5). Ecosystem nodulation and total N_2 fixation increased steeply to a maximum in our 12-year forests (29 $kg\,N\,ha^{-1}\,yr^{-1}$), and then decreased to low levels in 80-year forests. At 300 years, fixation recovered to 8% of maximum (Table 1, using scenario number 4 in Supplementary Note 3). Because our sampling was limited to specific forest ages, we cannot evaluate whether even higher fixation may have occurred between 5 and 30 years of forest age.

Highest nodulation and fixation coincided with the period of most rapid plant biomass accumulation (Fig. 1a), implying that nodulation and N_2 fixation were induced by high demand for nitrogen imposed by

[1]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA. [2]Smithsonian Tropical Research Institute, Balboa, Ancón, Panamá, Panama. [3]Forest Ecology and Forest Management Group, Centre for Ecosystem Studies, Wageningen University, 6700 AA Wageningen, The Netherlands. [4]Department of Geosciences and Natural Resource Management, University of Copenhagen, Rolighedsvej 23, DK-1958 Frederiksberg C, Denmark. [5]School of Forestry and Environmental Studies, Yale University, New Haven, Connecticut 06511, USA. †Present address: German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany.

Table 1 | Properties of N_2 fixers and fixation at the ecosystem, community and individual scales during succession

Variable	Forest age				
	5 years (4 plots)	12 years (4 plots)	30 years (4 plots)	80 years (2 plots)	300 years (2 plots)
Ecosystem: total N_2 fixation (kg N ha^{-1} yr^{-1})	10 (4)	29 (12)	11 (6)	0.3 (0.1)	2.2 (0.2)
Community: fixer basal area (per cent of total)	5.0 (1.7)	14.5 (1.1)	5.4 (2.3)	2.6 (0.9)	5.8 (1.4)
Community: fixer versus non-fixer recruitment (per cent per 2 years)	61 versus 5.2	0.51 versus −2.9	−1.9 versus −1.15	NA	NA
Individual: ratio of fixer to non-fixer growth rate	9.2 (3.7)	2.3 (0.6)	1.1 (0.9)	NA	NA

Numbers represent mean (s.e.m.). Ecosystem total N_2 fixation is calculated based on species- and forest-age-specific data for both nodulation and per-nodule biomass fixation (scenario number 4 in Supplementary Note 3). Recruitment was calculated as the per cent of trees remaining in 2011 compared to 2009. Positive numbers indicate increasing abundance (that is, positive recruitment), and negative numbers signify that the group is decreasing (that is, self-thinning). NA, numbers not available because the forest inventory was limited to 1 year. Growth rate was calculated as the average difference in per-individual biomass between 2009 and 2011. Values of greater than 1 indicate greater growth of fixers relative to non-fixers, whereas values of less than 1 indicate that fixers are growing slower relative to non-fixers.

the regrowing forest. In turn, fixation decreased between the 12-, 30- and 80-year-old forests, coincident with a progressive slowdown of biomass accumulation; fixation in 300-year-old forests was similar to independent measures[15] from nearby old-growth forests (red point, Fig. 1b).

These ecosystem trends in N_2 fixation were corroborated by independent observations at scales of plant community composition and individual plant growth (Table 1). At the community scale, relative fixer abundance (per cent of total basal area) displayed the same trend as ecosystem nodulation and fixation: a steep increase to a peak in 12-year forests, followed by a decrease to low levels in forests of 80 years and older. At the individual scale, fixers accumulated biomass ninefold faster than non-fixers at 5 years into forest recovery, but the fixer growth advantage declined to twofold in 12-year-old forests, and became equivalent to non-fixers in 30-year-old forests. This fixer growth advantage was also apparent at the community scale in that recruitment of fixers greatly exceeded non-fixers (61% versus 5%) in young forests, but decreased to similar levels in the progressively older forests. The partial recovery of fixation and percentage basal area of fixers in 300-year-old forests was due to increased abundance and per-individual fixation of *Tachigali versicolor* (discussed below; see also Supplementary Note 4).

We next asked how these emergent patterns in fixation depended upon the identity of N_2-fixing species. When assessed across all plots, individual species ranged approximately 200-fold in total nodulation and fixation (Fig. 2a and Table 2), with two species contributing >65% of all nodulation: *Inga cocleensis* and *T. versicolor*.

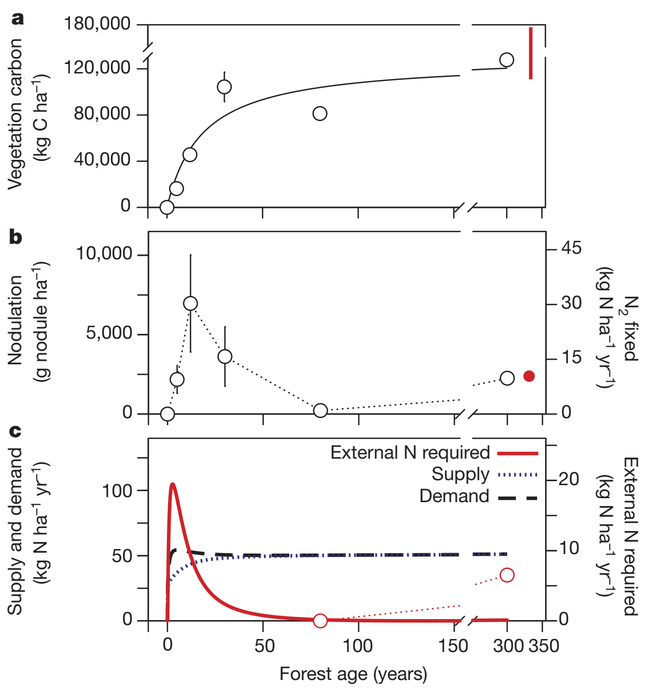


Figure 1 | Biomass carbon accumulation and dinitrogen fixation in forests recovering from land use. **a**, Sum of above- and belowground vegetation carbon (circles). Error bars (means ± s.e.m.) are invisible when replicates were similar. The fitted Michaelis–Menten curve was used to determine net nitrogen demand for the model, below. Red bar identifies the biomass range observed across 11 (ref. 25) old-growth Panamanian forests. **b**, Nodule biomass and N_2 fixed (mean ± s.e.m.). Fixation differs slightly from Table 1 because constant per-nodule biomass fixation rate is assumed (scenario number 1, Supplementary Note 3). Red point identifies fixation in nearby mature forests[15]. **c**, Modelled plant nitrogen demand (black), soil nitrogen supply (blue) and external nitrogen required from N_2 fixation (red). Dotted red line illustrates that fixation increases with disturbance in mature forests. $n = 4$ 0.2-ha plots for 5–30 years, $n = 2$ 1.0-ha plots for 80–300 years.

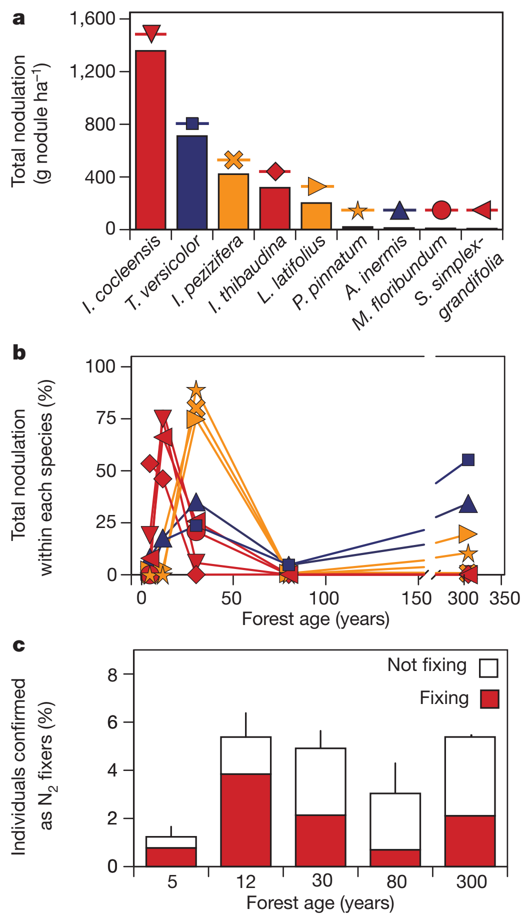


Figure 2 | Individual species contributions to ecosystem fixation. **a**, Calculated total nodulation by species: *Inga cocleensis* ($n = 64$), *Tachigali versicolor* ($n = 36$), *Inga pezizifera* ($n = 29$), *Inga thibaudiana* ($n = 30$), *Lonchocarpus latifolius* ($n = 21$), *Platymiscium pinnatum* ($n = 12$), *Andira inermis* ($n = 30$), *Machaerium floribundum* ($n = 11$) and *Swartzia simplex-grandifolia* ($n = 35$). **b**, Observed percentage of total fixation for each species at each forest age, and the species that specialize on young (red), mid-successional (orange) and mature (blue) forests (colours and symbols are same as in **a**). n values are given in Supplementary Data. **c**, Proportion of trees with known ability to fix that are currently fixing (red) and not fixing (white). $n = 4$ 0.2-ha plots for 5–30 years, $n = 2$ 1-ha plots for 80–300 years. Error bars indicate s.e.m.

Although six of the nine N_2-fixing species were present throughout the chronosequence, individual species differed in when they fixed the most and were most abundant (Fig. 2b and Table 2). One group dominated by *I. cocleensis* fixed primarily in the youngest forests, a second group in 30-year-old forests (dominated by *Inga pezizifera*), and a third group in mature forests (dominated by *T. versicolor*). We infer that these forests depend on a diverse suite of tree species that partition the biogeochemical niche[26] for fixation over the course of biomass recovery.

We further asked whether our observed fixation trend was generated by, first, N_2-fixing species that always fix but are progressively excluded from the community as nitrogen is replenished (that is, obligate fixation); or, second, N_2-fixing species that persist over time but that turn fixation on or off depending on their nitrogen balance (that is, facultative fixation). This distinction is significant because the first theory implies that the ecosystem capacity for fixation disappears over time, whereas the second theory implies that the capacity is functionally ensured throughout forest succession.

Our results support the second theory. The proportion of trees with confirmed ability to fix remained relatively stable in forests 12 years and older, yet the proportion that expressed fixation declined from 71% at 12 years to 23% at 80 years (Fig. 2c and Table 2). Furthermore, even when fixing, individuals varied their investment in fixation and nodulation across at least two orders of magnitude independent of species (Table 2), consistent with facultative fixation at the individual scale. These dynamics indicate that fixers can buffer forest nitrogen supply by upregulating fixation when nitrogen is low[15,27].

Our findings identify an ecosystem-scale feedback between forest regrowth, vegetation nitrogen demand and the supply of new nitrogen ensured by N_2 fixers. The results imply that nitrogen limitation is the mechanism favouring fixers and fixation in young forests, but the drop in fixation in older forests could result from either an alleviation of nitrogen-limitation or a decrease in light available to fixers. However, light does not explain the approximately 99% decline between maximum and minimum fixation rates (12 versus 80 years), because the proportion of fixers with canopy access to sunlight remained high over time (51% versus 45%, respectively; Supplementary Note 8 and Supplementary Table 7).

We evaluated the nitrogen feedback using a mass-balance model that captures the mechanisms of plant nitrogen demand (net leaf, wood and fine-root growth), plant–soil nitrogen recycling, nitrogen inputs (deposition and fixation) and nitrogen losses (Methods, Supplementary Notes 5–7, Supplementary Box 1, Supplementary Table 6 and Supplementary Figs 3, 6 and 7). When the model was run with the observed pattern of biomass carbon and nitrogen accumulation (fitted line in Fig. 1a), it revealed, first, that the nitrogen demand imposed by vegetation growth greatly exceeded the capacity of internal recycling to supply plant nitrogen in the first approximately 25 years (dashed versus dotted lines, Fig. 1c); second, that the resulting nitrogen deficit (red line, Fig. 1c) was highest when biomass accumulation was steepest in early succession, and declined as biomass accumulation decelerated over time (red line in Fig. 1c versus black line in Fig. 1a); and, third, that the overall shape of the resulting nitrogen deficit recreated the observed changes in fixation across our plots (red line in Fig. 1c versus black line in Fig. 1b). The field data indicated a lag in fixation in our youngest forests, compared to the model, probably caused by slow establishment and growth of the early tree community (Supplementary Note 6 and Supplementary Figs 4 and 5).

Although fixation would be downregulated as the plant–soil nitrogen cycle becomes increasingly balanced (Fig. 1c), we nevertheless observed low but continued fixation in 300-year-old forests by the gap specialist, *T. versicolor*. This indicates that the feedback between

Table 2 | Properties of individual species at different stages of forest recovery

Forest age (years)	Species (from left to right in decreasing order of importance for N_2 fixation at the ecosystem level)								
	I. cocleensis	*T. versicolor*	*I. pezizifera*	*I. thibaudiana*	*L. latifolius*	*P. pinnatum*	*A. inermis*	*M. floribundum*	*S. simplex-grandifolia*
Abundance, given as the number of trees per species per ha; the total number of trees of any species per ha is given in parentheses beside forest age									
5 (13,806)	230	0	10	135	65	0	95	0	30
12 (6,145)	460	5	80	150	235	0	25	7	140
30 (3,925)	70	95	100	15	5	60	70	8	165
80 (2,240)	8	31	22	7	5	3	4	0	5
300 (1,569)	9	48	12	6	18	6	24	0	0
Trees that were N_2-fixing (per cent of trees examined)									
5	62	NA	NA	82	83	NA	17	NA	100
12	91	100	81	87	60	NA	50	22	40
30	56	53	54	100	100	72	17	50	25
80	0	48	33	NA	25	0	17	NA	0
300	8	72	25	NA	45	50	30	NA	0
Mean nodulation rate for N_2-fixing trees (min–max) (g nodule m^{-2} under tree canopy)									
5	1.02 (0.06–9.34)	NA	NA	1.07 (0.11–6.66)	1.17 (0.04–3.00)	NA	0.18	NA	0.82
12	1.93 (0.15–7.79)	8.12	1.35 (0.2–12.87)	1.43 (0.30–8.40)	0.34 (0.18–0.49)	NA	0.15	1.54 (0.15–16.01)	0.63 (0.08–3.66)
30	1.81 (0.02–4.78)	6.73 (0.30–31.87)	3.09 (0.25–8.03)	1.18 (0.3–3.78)	9.05	1.08 (0.27–2.26)	0.27 (0.21–0.36)	4.94	0.55 (0.11–3.23)
80	NA	1.82 (0.44–2.95)	0.79	NA	0.19	NA	0.11	NA	NA
300	3.33	3.06 (0.18–16.01)	0.46	NA	0.99 (0.32–5.62)	1.18 (0.14–9.96)	0.18 (0.06–0.39)	NA	NA
Mean fixation rate for N_2-fixing trees* (s.e.m.) (g N m^{-2} yr^{-1} under tree canopy)									
5	4.6 (2.2)	NA	NA	4.8 (2.1)	0.8 (0.8)	NA	0.5 (NA)	NA	2.2 (NA)
12	8.7 (2.4)	7.4 (NA)	6.5 (2.8)	6.5 (2.8)	0.2 (0.1)	NA	0.4 (NA)	4.1 (NA)	1.7 (0.3)
30	8.2 (4.8)	6.1 (3.6)	5.3 (NA)	5.3 (NA)	6.4 (NA)	2.9 (1.4)	0.7 (NA)	13.2 (NA)	1.5 (1.2)
80	NA	1.7 (0.3)	2.6 (NA)	NA	0.1 (NA)	NA	0.3 (NA)	NA	NA
300	11.1 (NA)	2.8 (1.1)	1.6 (NA)	NA	0.7 (0.2)	3.1 (NA)	0.5 (NA)	NA	NA

* Uses the species- and forest-age-specific per-nodule biomass fixation rate. NA, data not available because the species was either absent or not measured, or because the error could not be calculated due to insufficient replication. The number of trees sampled for Table 2 is given in the Supplementary Data for Fig. 2b.

fixation and soil nitrogen deficiency also can occur in tree-fall gaps[15] that begin to form as stands mature and become less uniform; *T. versicolor* grows rapidly when such light gaps form.

Our findings offer an example in which a vital ecosystem function emerges from the properties of individual species. We estimate that fixers provided >50% of the nitrogen needed to support the sequestration of 50,000 kg of plant carbon per hectare over the first 12 years of forest recovery. The presence of such high fixation rates on the phosphorus-poor Agua Salud soils implies that phosphorus may be less of a constraint on tropical fixation than often assumed in theories or models. Other factors including competition for light or soil molybdenum[28] may ultimately limit fixers and/or fixation. Most fundamentally, our results offer a mechanism by which tropical forests can avoid nitrogen constraints caused by disturbance or land-use transition, or that are predicted to occur in response to rising atmospheric CO_2.

METHODS SUMMARY

We sampled 16 forest plots at different stages of recovery from agriculture to intact forest within the 15 km^2 Agua Salud Project and in the nearby Soberania National Park. All plots had similar climate and were on soils (oxisols) with similar chemical and physical characteristics. In each plot, we measured the diameter of trees and calculated aboveground and belowground carbon and nitrogen biomass using allometric equations developed from 244 trees collected at our sites. We analysed leaf carbon and nitrogen contents from 55 common species across our plots, and used standard conversion factors to estimate carbon and nitrogen biomass across leaf, roots and wood tissues of each individual tree. To determine whole-ecosystem biomass carbon and nitrogen, we summed across all individuals of each species in each plot.

We quantified fixation in individuals and at the ecosystem scale by measuring nodule biomass for 268 trees across the 9 most abundant N_2-fixing species, and per-nodule biomass fixation rates derived from ^{15}N incorporation and acetylene reduction. We calculated fixation at the scale of the ecosystem across both individual species (sum of all individuals of a given species) and across all species combined (sum of all individuals from all N_2-fixing species).

We developed a mass-balance plant–soil ecosystem model in which plant nitrogen demand and plant litter input to soils was externally forced by the observed carbon accumulation curve across our chronosequence. We considered a plant-available and a plant-unavailable soil pool, and a 'passive' soil pool that turned over on a timescale greater than the successional dynamics of this study (and thus was considered external to the plant–soil model). Nitrogen can enter the system through inputs from deposition, the slow mineralization of the externalized passive pool or N_2 fixation, which is determined as the difference between plant nitrogen demand and soil nitrogen supply.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 November 2012; accepted 2 August 2013.
Published online 15 September 2013.

1. Davidson, E. A. *et al.* Recuperation of nitrogen cycling in Amazonian forests following agricultural abandonment. *Nature* **447,** 995–998 (2007).
2. Zaehle, S. & Dalmonech, D. Carbon–nitrogen interactions on land at global scales: Current understanding in modeling climate biosphere feedbacks. *Curr. Opin. Environ. Sustain.* **3,** 311–320 (2011).
3. Gerber, S., Hedin, L. O., Oppenheimer, M., Pacala, S. W. & Shevliakova, E. Nitrogen cycling and feedbacks in a global dynamic land model. *Glob. Biogeochem. Cycles* **24,** GB1001 (2010).
4. Thornton, P. E., Lamarque, J.-F., Rosenbloom, N. A. & Mahowald, N. M. Influence of carbon–nitrogen cycle coupling on land model response to CO_2 fertilization and climate variability. *Glob. Biogeochem. Cycles* **21,** GB4108 (2007).
5. Houghton, R. A., Hall, F. & Goetz, S. J. Importance of biomass in the global carbon cycle. *J. Geophys. Res.* **114,** G00E03 (2009).
6. Hungate, B. A., Dukes, J. S., Shaw, R., Luo, Y. & Field, C. Nitrogen and climate change. *Science* **302,** 1512–1513 (2003).
7. Sprent, J. I. *Legume Nodulation: A Global Perspective*. (Wiley-Blackwell, 2009).
8. Goll, D. S. *et al.* Nutrient limitation reduces land carbon uptake in simulations with a model of combined carbon, nitrogen and phosphorus cycling. *Biogeosci. Discuss.* **9,** 3173–3232 (2012).
9. Brown, S. & Lugo, A. E. Tropical secondary forests. *J. Trop. Ecol.* **6,** 1–32 (1990).
10. Pan, Y. *et al.* A large and persistent carbon sink in the world's forests. *Science* **333,** 988–993 (2011).
11. Davidson, E. A. *et al.* Nitrogen and phosphorus limitation of biomass growth in a tropical secondary forest. *Ecol. Appl.* **14,** S150–S163 (2004).
12. Amazonas, N. T., Martinelli, L. A., Piccolo, M. C. & Rodrigues, R. R. Nitrogen dynamics during ecosystem development in tropical forest restoration. *For. Ecol. Manage.* **262,** 1551–1557 (2011).
13. Russell, A. E. & Raich, J. W. Rapidly growing tropical trees mobilize remarkable amounts of nitrogen, in ways that differ surprisingly among species. *Proc. Natl Acad. Sci. USA* **109,** 10398–10402 (2012).
14. Gehring, C., Vlek, P. L. G., de Souza, L. A. G. & Denich, M. Biological nitrogen fixation in secondary regrowth and mature rainforest of central Amazonia. *Agric. Ecosyst. Environ.* **111,** 237–252 (2005).
15. Barron, A. R., Purves, D. W. & Hedin, L. O. Facultative nitrogen fixation by canopy legumes in a lowland tropical forest. *Oecologia* **165,** 511–520 (2011).
16. Binkley, D., Senock, R. & Cromack, K. Jr. Phosphorus limitation on nitrogen fixation by *Facaltaria* seedlings. *For. Ecol. Manage.* **186,** 171–176 (2003).
17. Vitousek, P. M. & Howarth, R. W. Nitrogen limitation on land and in the sea: how can it occur? *Biogeochemistry* **13,** 87–115 (1991).
18. van Groenigen, K. J. *et al.* Element interactions limit soil carbon storage. *Proc. Natl Acad. Sci. USA* **103,** 6571–6574 (2006).
19. Hedin, L. O., Brookshire, E. N. J., Menge, D. N. L. & Barron, A. R. The nitrogen paradox in tropical forest ecosystems. *Annu. Rev. Ecol. Evol. Syst.* **40,** 613–635 (2009).
20. Isbell, F. *et al.* High plant diversity is needed to maintain ecosystem services. *Nature* **477,** 199–202 (2011).
21. Yang, X., Richardson, T. K. & Jain, A. K. Contributions of secondary forest and nitrogen dynamics to terrestrial carbon uptake. *Biogeosciences* **7,** 3041–3050 (2010).
22. van Breugel, M., Ransijn, J., Craven, D., Bongers, F. & Hall, J. S. Estimating carbon stock in secondary forests: decisions and uncertainties associated with allometric biomass models. *For. Ecol. Manage.* **262,** 1648–1657 (2011).
23. Quesada, C. A. *et al.* Variations in chemical and physical properties of Amazon forest soils in relation to their genesis. *Biogeosciences* **7,** 1515–1541 (2010).
24. Fyllas, N. M. *et al.* Basin-wide variations in foliar properties of Amazonian forest: phylogeny, soils and climate. *Biogeosci. Discuss.* **6,** 3707–3769 (2009).
25. Chave, J. *et al.* Error propagation and scaling for tropical forest biomass estimates. *Phil. Trans. R. Soc. Lond. B* **359,** 409–420 (2004).
26. Menge, D. N. & Hedin, L. O. Nitrogen fixation in different biogeochemical niches along a 120000-year chronosequence in New Zealand. *Ecology* **90,** 2190–2201 (2009).
27. Batterman, S. A., Wurzburger, N. & Hedin, L. O. Nitrogen and phosphorus interact to control tropical symbiotic N_2 fixation: a test in *Inga punctata*. *J. Ecology* (in the press).
28. Barron, A. R. *et al.* Molybdenum limitation of asymbiotic nitrogen fixation in tropical forest soils. *Nature Geosci.* **2,** 42–45 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Adelberg and K. Zelazny for assisting with data collection, M. Baillon and A. Hernandez for botanical identifications, N. Wurzburger and A. Barron for species-specific N_2 fixation rates, J. Sprent for advice about N_2-fixing trees and P. Reich for comments. This work was supported by grants to L.O.H. from the National Science Foundation (NSF; DEB-0614116), the National Oceanic and Atmospheric Association (NOAA; grant NA17RJ262 – 344), the Cooperative Institute for Climate Science of Princeton University and the Carbon Mitigation Initiative of Princeton University; and to S.A.B. from the Smithsonian Tropical Research Institute (STRI). It is a contribution to the Agua Salud Project (ASP), a collaboration among STRI, the Panama Canal Authority (ACP) and the National Environmental Authority of Panama (ANAM). ASP funding came from the HSBC climate partnership, STRI, the Frank Levinson Family Foundation, the Motta Family Foundation and an anonymous donor.

Author Contributions S.A.B., L.O.H., J.S.H. and M.v.B. designed the project. S.A.B. conducted N_2-fixation-related field work. J.R., M.v.B. and J.S.H. provided allometry data; D.J.C. provided plant nutrient data. S.A.B. and L.O.H. wrote the paper. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.A.B. (sbatterm@princeton.edu).

METHODS

Study sites. We sampled 16 forest plots at different stages of recovery from agriculture to intact forest within the 15-km^2 Agua Salud Project[22] and in Soberania National Park approximately 3 km from Agua Salud. Soils across all sites were highly phosphorus-poor and clay-rich oxisols derived from basalt parent material, with no significant differences in carbon, nitrogen, phosphorus or cations across forest age ($P > 0.05$, ANOVA, d.f. = 4 for total carbon, total nitrogen, carbon:nitrogen, total phosphorus; and d.f. = 2 for Mg, K, Ca, Al and extractable phosphorus) (Supplementary Table 1). All sites were classified as tropical moist forest, received similar annual rainfall (2,700 mm; S. Paton, personal communication) with a December-to-April dry season, and shared a diverse community of tree species (>297). Forests contained 24 N_2-fixing tree taxa, many of which are commonly found throughout neotropical forests. We measured all biological and soil properties in the wet season.

Sampling approach. We selected sites to create a replicated sequence of forest recovery ages, with all sites <80 years old recovering from grazing, and sites 80 and 300 years old recovering from either cutting alone or cutting followed by grazing and/or non-intensive agriculture[29,30]. We used a standard chronosequence approach[31] in which space-for-time substitution is used to infer changes in ecosystem properties at temporal scales that exceed the limit of direct observation, and which assumes that sites share similar initial conditions and similar dynamics over time. Our forests and soils are very similar in composition and historical origin, but, depending on forest response to CO_2 fertilization and/or temperature change (at present unknown), increased atmospheric CO_2 or temperature may have introduced a historic bias in the rate of recovery in younger forest stands.

We adjusted area and number of plots to account for differences in tree density (13,800 versus 1,500 trees per ha in 5- versus 300-year-old forests, respectively; Table 2), landscape variability in community composition (high in young forests), biomass accumulation (high in young forests; Fig. 1a) and the influence of localized forest gap dynamics (absent in young forests but common in 300-year-old forests). Note that in mature forest it is not possible to conceptually or visually separate 'gaps' (unless they are only a few years old) from 'mature forest'. We sampled four replicate 0.2-ha plots in all forests <80 years, and two replicate 1.0-ha plots in the 80- and 300-year-old forests. In plots <80 years, we measured every tree with a diameter at breast height (DBH) of ≥5 cm, and half of all trees with a DBH of 1 to 5 cm; in 80- and 300-year-old plots we measured every tree with a DBH of ≥10 cm and two-fifths of trees with a DBH of 1 to 10 cm in 2008. We sampled all plots of <80 years in age in 2009 and 2011; this enabled us to reliably calculate differences in individual tree growth rates between fixers and non-fixers, and to quantify changes in community composition during the period of highest biomass accumulation and strongest competition for light. We did not re-sample our older sites because growth calculated from DBH becomes reliable only in large trees over longer timescales.

Biomass calculations. We used a standard allometric approach to quantify biomass allocation to leaves, wood and fine roots across all plots (Supplementary Fig. 1 and Supplementary Tables 2 and 3). For each of our 26 most common species (contributing approximately 75% of total basal area across all plots) we developed equations (all $R^2 > 0.85$) for aboveground wood and leaf biomass by collecting 244 trees <26 cm DBH: $B_{x,s} = Y_{x,s} \times \mathrm{DBH}^{(Z_{x,s})}$, where B is biomass, Y is a proportionality coefficient, Z is a scaling exponent, x is leaf or wood, and s is species[22] (Supplementary Table 2). For all other species, we used a general equation based on all species combined (Supplementary Table 2). Roots and the sum of aboveground and belowground wood were calculated as described below. For trees with a DBH of ≥26 cm, we used a general equation for aboveground tropical forest biomass[25] (AGB): $\mathrm{AGB} = A \times e^B$, where $B = -3.742 + 3.45 \times \log(\mathrm{DBH}) - 0.148 \times \log(\mathrm{DBH})^2$ and A is a coefficient for wood density. We calculated biomass allocation to leaves versus roots versus wood, as follows: leaf area index equilibrating at approximately 0.2 kg N m^{-2} after canopy closure; fine roots equal to leaf biomass[9,32]; coarse roots from a published allometric relation with stem biomass derived from the global Luyssaert database[33]; and total wood = AGB − leaf biomass + coarse root biomass. Biomass was summed across all trees in each plot and converted to carbon based on the average carbon content (0.47) in our forests[34].

Tissue nutrient pools. We determined biomass, and biomass carbon and nitrogen pools (Supplementary Table 3), by summing the biomass of individual trees (estimated with allometric equations as described above) across all individuals and species and applying the nitrogen content of tissues that we determined from our forests and the literature. We measured species-specific leaf nitrogen content of leaves from the 55 most common species of trees across the Agua Salud forest sites[35]. Leaf nitrogen contents were within the range found at other neotropical sites[24,36]. To calculate the total leaf nitrogen pool size, we found the community abundance-weighted average leaf nitrogen content and applied that to our total leaf biomass pool size. We used a standard nitrogen content of wood derived from the literature[37] in the lower (and conservative, for our calculations) range of tropical wood nitrogen contents[38], and we approximated fine-root nitrogen content as being equal to leaves. Total forest biomass nitrogen was determined as the sum of leaf, wood, and root nitrogen.

Dinitrogen fixation. We directly quantified nodule biomass and N_2 fixation by collecting 2,334 soil cores (5.5 cm in diameter, 10 cm deep) randomly from the rooting zone beneath each individual tree using crown area to approximate root area in the rainy season of 2010 (refs 15, 39). To account for differences in rooting area with tree size, we collected 6 cores for trees with DBH of 1–5 cm, 8 cores for trees with DBH of 5–10 cm, 10 cores for trees with DBH of 10–20 cm and 12 cores for trees with DBH >20 cm, consistent with our sampling method that we developed over the past 9 years. All nodules were cut open to confirm nitrogenase activity, and total nodule biomass was expressed per soil area of each tree (g nodule m^{-2}) and per individual tree (g nodule tree^{-1}). Because natural abundance $^{15}N/^{14}N$ ratios cannot necessarily resolve individual tree N_2 fixation[40], we quantified nodulation from June to August 2010 and nodule mass-specific fixation using ^{15}N-tracer incubations of several species of *Inga*[15] and acetylene reduction assay (ARA) measures in 15 species of fixers across different growth conditions in Panamanian forests between 2003 and 2012 (Supplementary Table 4). We expanded these measures to the plot scale as follows: we scaled average species-specific nodule biomass (g nodule m^{-2}) to the rooting area for each individual tree and summed across all trees in each plot; for the species for which we did not directly quantify nodulation (<30% of fixer basal area), we used the average nodule biomass per area from all species. When evaluating percent nodulation across species (Fig. 2c) and the fraction of trees that were currently fixing, we considered only trees for which we had directly measured fixation.

We calculated N_2 fixation by multiplying nodule biomass by per-nodule biomass fixation rate from the two groups above, and evaluated the potential effect of species- and forest-age-specific per-nodule biomass fixation rates (discussed in Supplementary Note 3). Total ecosystem fixation rates were relatively insensitive to variations in these assumptions; we therefore report species- and forest-age-specific data on both nodulation and per-nodule biomass fixation in Table 1 and compare the alternative scenarios in Supplementary Note 3.

Model description. Our model evaluates the effect of the observed biomass accumulation trend (line in Fig. 1a; quantified from allometric measures of all individual trees across all species and all plots) on the internal forest N cycle. The model (Supplementary Box 1) tracks nitrogen mass balances (as kg N ha^{-1}) in plant pools of leaves (N_L), wood (N_W) and fine roots (N_R), as well as in plant available (N_A) and plant unavailable (N_U) forms in soil:

$$B_x = B^*_x\, t/(t^*_x + t) \qquad (1)$$

$$N_L = N_R = \omega_L B_L \qquad (2)$$

$$N_W = \omega_W\,(B_T - B_L - B_R) \qquad (3)$$

$$dN_A/dt = \mathrm{I} - \text{plant N demand} + mN_U - k\,(N_A) + r\,\alpha\,(\mu_L N_L(t) + \mu_R N_R(t)) \qquad (4)$$

$$dN_U/dt = \beta\mu_W N_W(t) - m(N_U) + r\,(1-\alpha)\,(\mu_L N_L(t) + \mu_R N_R(t)) \qquad (5)$$

Equation (1) describes the observed net plant biomass accumulation (B) over time (t) fitted as a Michaelis–Menten equation based on the tissue-specific allometric calculations outlined above (Fig. 1a and Supplementary Note 2), and where B^*_x is the equilibrium maximum biomass, t^*_x is the half-saturation constant and x refers to tissue pool (L, leaves; R, fine roots; W, wood; T, total biomass). Following standard convention[9,32,33], we assumed equal quantities of nitrogen in leaf (N_L) and fine root (N_R) biomass, and calculated both pools as leaf biomass times the fraction of nitrogen in leaf biomass (ω_L) (equation (2)). We calculated nitrogen in wood as the product of wood biomass (total plant biomass minus leaf and fine-root biomass) and the fraction of nitrogen in wood (ω_w) (equation (3)). Equation (4) gives the time-dependent change in the plant-available soil nitrogen pool, as a function of external nitrogen input (I), the plant nitrogen growth demand, the input mineralization rate (m) from plant-unavailable soil nitrogen, the loss rate (k) from the plant-available soil nitrogen pool, and the input rates of plant leaf (μ_L) and fine-root (μ_R) turnover, where r is the nitrogen fraction not retranslocated and α is the

fraction entering the soil-available pool. Plant nitrogen demand was calculated from equation (1) applied to all tissue pools and corrected for annual turnover (adjusted for retranslocation) of leaf and fine-root nitrogen. Equation (5) gives the time-dependent change in the plant-unavailable soil nitrogen pool, as a function of the input rate from wood mortality (μ_W), where β identifies the fraction of wood mortality entering the plant-unavailable pool; the loss mineralization rate (m) to the plant-available pool; and the input rates of nitrogen from leaf (μ_L) and fine-root (μ_R) turnover.

We simplified our model by externalizing all soil nitrogen with mean residence time (MRT) that exceeded the dynamics of interest in this study (taken as $>$30 years). We assumed that this soil nitrogen is large in size and originated from wood, with the flux calculated as a loss fraction ($1 - \beta$) of wood mortality. We also considered a slow but constant mineralization input from this external pool, which we added to the external input term (I). We show in Supplementary Information that this simplification offered a good approximation of more complex models in which we explicitly tracked a passive soil nitrogen pool following the standard approach of CENTURY-type models.

The external nitrogen required (that is, N_2 fixation) was calculated as the difference between plant nitrogen demand and soil nitrogen supply. A positive value indicates that the internal nitrogen supply could not match the demands of the recovering forest, thus necessitating additional nitrogen inputs from symbiotic fixation. A negative balance, in turn, indicates that internal supplies were sufficient or exceeded forest demand.

Model parameterization. We parameterized the model using values reasonable for our Panamanian forests as summarized in Supplementary Table 6. We used the tissue nitrogen contents (percentage nitrogen) that we determined from our data and the literature as described above. We allowed leaves to turn over annually, which is conservative relative to our observations across the youngest forests, as faster turnover would increase external nitrogen required.

We doubled the rate of wood turnover (from 0.02 to 0.04 per year) to evaluate the effect of disturbance on fixation in mature forests (shown as dotted red line in Fig. 1c). Because we are simulating forest carbon recovery during secondary succession, we started the plant-unavailable nitrogen pool at 60% of equilibrium size; we show in Supplementary Information that increasing the pool size to 100% does not alter the dynamics of the system.

29. Hassler, S. K., Zimmerman, B., van Breugel, M., Hall, J. S. & Elsenbeer, H. Recovery of saturated hydraulic conductivity under secondary succession on former pasture in the humid tropics. *For. Ecol. Manage.* **261,** 1634–1642 (2011).
30. Neumann-Cosel, L., Zimmerman, B., Hall, J. S., van Breugel, M. & Elsenbeer, H. Soil carbon dynamics under young tropical secondary forests on former pastures—a case study from Panama. *For. Ecol. Manage.* **261,** 1625–1633 (2011).
31. Walker, L. R., Wardle, D. A., Bardgett, R. D. & Clarkson, B. D. The use of chronosequences in studies of ecological succession and soil development. *J. Ecol.* **98,** 725–736 (2010).
32. Jackson, R. B., Mooney, H. A. & Schulze, E. D. A global budget for fine root biomass, surface area, and nutrient contents. *Proc. Natl Acad. Sci. USA* **94,** 7362–7366 (1997).
33. Wolf, A., Field, C. & Berry, J. A. Allometric growth and allocation in forests: a perspective from FLUXNET. *Ecol. Appl.* **21,** 1546–1556 (2011).
34. Martin, A. R. & Thomas, S. C. A reassessment of carbon content in tropical trees. *PLoS ONE* **6,** e23533 (2011).
35. Craven, D. J. *Dynamics of Tropical Secondary Forests in Central Panama: Linking Functional Traits with Ecological Performance during Succession.* PhD thesis, Yale Univ. (2012).
36. Townsend, A. R., Cleveland, C. C., Asner, G. P. & Bustamente, M. M. C. Controls over foliar N:P ratios in tropical rain forests. *Ecology* **88,** 107–118 (2007).
37. Martius, C. Density, humidity, and nitrogen content of dominant wood species of floodplain forests (várzea) in Amazonia. *Eur. J. Wood Wood Products* **50,** 300–303 (1992).
38. Chave, J. *et al.* Towards a worldwide wood economics spectrum. *Ecol. Lett.* **12,** 351–366 (2009).
39. Muller-Landau, H. C. *et al.* Testing metabolic ecology theory for allometric scaling of tree size, growth and mortality in tropical forests. *Ecol. Lett.* **9,** 575–588 (2006).
40. Barron, A. R. *Patterns and Controls of Nitrogen Fixation in a Lowland Tropical Forest, Panama.* PhD thesis, Princeton Univ. (2007).

LETTER

doi:10.1038/nature12574

Bottom-up control of geomagnetic secular variation by the Earth's inner core

Julien Aubert[1], Christopher C. Finlay[2] & Alexandre Fournier[3]

Temporal changes in the Earth's magnetic field, known as geomagnetic secular variation, occur most prominently at low latitudes in the Atlantic hemisphere[1,2] (that is, from −90 degrees east to 90 degrees east), whereas in the Pacific hemisphere there is comparatively little activity. This is a consequence of the geographical localization of intense, westward drifting, equatorial magnetic flux patches at the core surface[3]. Despite successes in explaining the morphology of the geomagnetic field[4], numerical models of the geodynamo have so far failed to account systematically for this striking pattern of geomagnetic secular variation. Here we show that it can be reproduced provided that two mechanisms relying on the inner core are jointly considered. First, gravitational coupling[5] aligns the inner core with the mantle, forcing the flow of liquid metal in the outer core into a giant, westward drifting, sheet-like gyre[6]. The resulting shear concentrates azimuthal magnetic flux at low latitudes close to the core–mantle boundary, where it is expelled by core convection and subsequently transported westward. Second, differential inner-core growth[7,8], fastest below Indonesia[6,9], causes an asymmetric buoyancy release in the outer core which in turn distorts the gyre, forcing it to become eccentric, in agreement with recent core flow inversions[6,10,11]. This bottom-up heterogeneous driving of core convection dominates top-down driving from mantle thermal heterogeneities, and localizes magnetic variations in a longitudinal sector centred beneath the Atlantic, where the eccentric gyre reaches the core surface. To match the observed pattern of geomagnetic secular variation, the solid material forming the inner core must now be in a state of differential growth rather than one of growth and melting induced by convective translation[7,8].

The peculiar geographical localization (both in latitude and longitude, Fig. 1a and b, respectively) of geomagnetic secular variation observed during the historical era[1] (AD 1590–1990) has now been mapped to an unprecedented level of accuracy by virtue of more than one decade of global and continuous satellite magnetic observations. Recent models[12–14] benefit from improved removal of the contribution from external fields, and are now reliably able to resolve structures down to a 1,500-km lateral extent at the Earth's core–mantle boundary (spherical harmonic degree 13, Fig. 2a, b). To exploit this wealth of data fully, and uncover the dynamical processes operating in the Earth's core, it is necessary to combine magnetic observations with numerical models of the geodynamo[15]. These have steadily improved over the past two decades, with the standard models now well understood, thanks to concerted benchmarking activities and systematic exploration of the accessible parameter space[16]. Despite this progress, success in understanding the geomagnetic secular variation and predicting its future evolution remains conditional on our ability to explain its detailed structure with physical mechanisms simple and robust enough to withstand extrapolation from the simplified working regime of current models to the conditions of Earth's core. An outstanding difficulty for standard models has been how to reproduce the westward drift of low-latitude magnetic flux patches at the core–mantle boundary[3,14] in a self-consistent fashion. Quasi-steady westward drift has previously been obtained[17–19], but it becomes erratic[16] in models where magnetic field advection becomes realistically strong relative to diffusion. Furthermore, the Atlantic–Pacific asymmetry in the kinematics of these patches suggests some form of heterogeneous boundary control[6]. The combined effect on secular variation of top-down forcing from lower-mantle thermal anomalies[20] versus bottom-up forcing from the recently discovered inner-core translational instability[7,8] has remained unexplored.

Our coupled Earth numerical dynamo model (Methods) addresses these issues by building on the well understood basis of standard geodynamo models[16] and including some simple ingredients from early

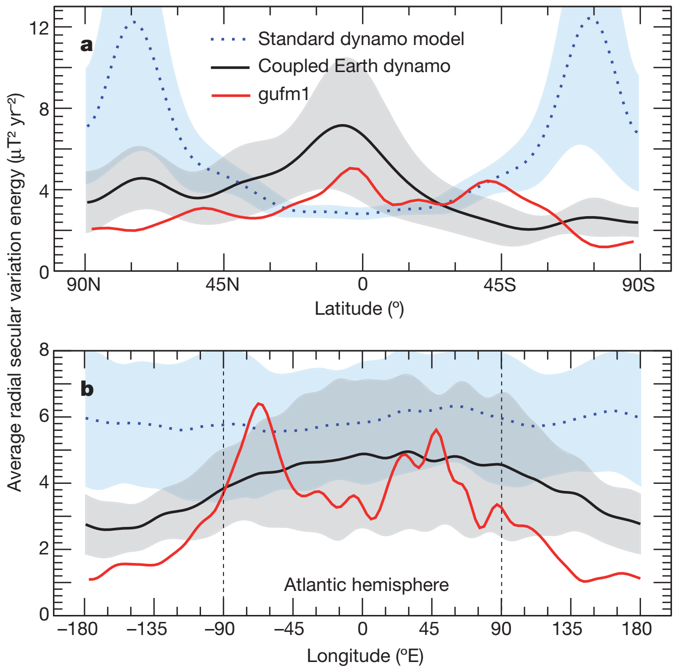


Figure 1 | Geographical localization of geomagnetic secular variation. **a**, **b**, Profiles of the time-averaged radial secular variation energy, averaged over longitude (**a**) or latitude (**b**). In each panel, data are taken from the historical geomagnetic field model[1] gufm1 (red line; spanning AD 1590–1990 and evaluated at the core–mantle boundary), from 400-year time-averages successively taken within a 83,000-year sequence of a standard numerical dynamo model (dotted blue line, average profile; blue shading, ±1 s.d. in the distribution of 400-year time averages), and within a 67,000-year sequence of our coupled Earth dynamo model (black line, average profile; grey shading, ±1 s.d. in the distribution of 400-year time averages). The numerical dynamo outputs are filtered at spherical harmonic degree and order 8, and the gufm1 model is presented at native resolution. The standard model has rigid, electrically insulating and non-moving boundaries. The coupled Earth dynamo model includes visco-magnetic outer-core/inner-core coupling, gravitational inner-core/mantle coupling, and thermochemical, heterogeneous coupling between the outer core, inner core and mantle.

[1]Dynamique des Fluides Géologiques, Institut de Physique du Globe de Paris, Sorbonne Paris Cité, Université Paris Diderot, UMR 7154 CNRS, F-75005 Paris, France. [2]Division of Geomagnetism, National Space Institute, Technical University of Denmark, Elektrovej, 2800 Kongens Lyngby, Denmark. [3]Géomagnétisme, Institut de Physique du Globe de Paris, Sorbonne Paris Cité, Université Paris Diderot, UMR 7154 CNRS, F-75005 Paris, France.

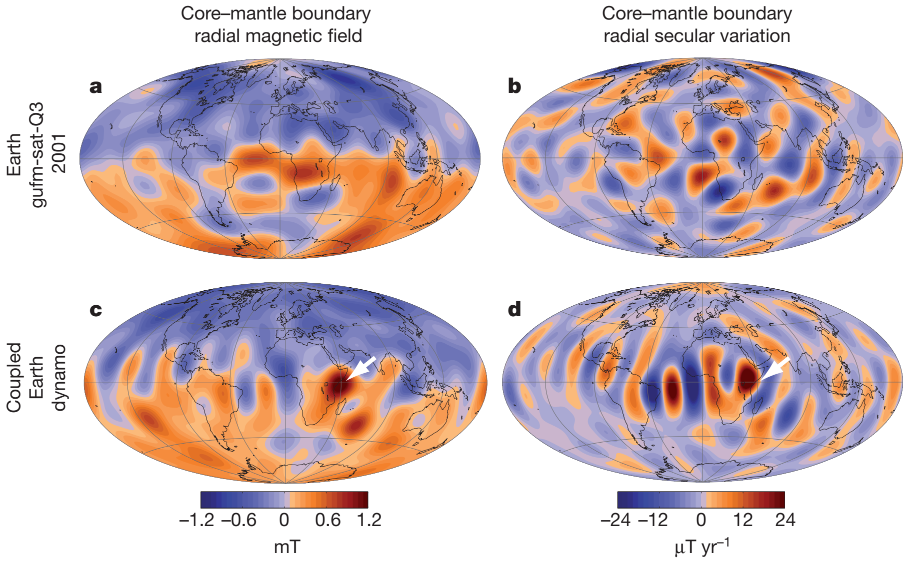


Figure 2 | Maps of the magnetic field and secular variation. a–d, Atlantic-centred Hammer projections of the core–mantle boundary radial magnetic field (**a**, **c**) and its temporal rate of change or radial secular variation (**b**, **d**); data are from geomagnetic field model[14] gufm-sat-Q3 in 2001 (**a**, **b**) and a snapshot of the coupled Earth dynamo model (**c**, **d**; orange denotes an outward magnetic field), both filtered at spherical harmonic degree and order 13. White arrows in **c** and **d** mark a location chosen for further analysis in Figs 3 and 4.

modelling efforts[17,21,22] that have recently been neglected. A field morphology similar to that of the Earth[4] (Fig. 2 and Extended Data Fig. 1) is achieved by choosing ratios of the timescales for magnetic induction, magnetic diffusion and the length of the day as close as computationally feasible to their Earth counterparts. We compromise on the ratio of the length of the day to the viscous diffusion timescale (the Ekman number, Ek), which, although low, lies an order of magnitude above state-of-the art values[18,19], in order to facilitate a parameter space exploration (Supplementary Information and Extended Data Table 1). The influence of viscosity is further mitigated by adopting a stress-free mechanical condition at the core–mantle boundary.

Robust westward drift and time-dependence of low-latitude secular variation similar to that observed over the past 400 years are achieved in the coupled Earth dynamo model through a mechanism based on indirect exchanges of angular momentum between the outer core and the mantle via the inner core. The inner core couples with the outer core primarily through a magnetic torque, which dominates at the Ekman number studied. At the base of the outer core, thermochemical 'winds' resulting from the interaction of the Coriolis force with convective upwellings entrain the inner core in the eastward direction[21,23] with respect to the planetary rotating frame. But a gravitational torque[5,22–24], resulting from coupling between lower-mantle mass anomalies and the induced topography at the inner-core surface, transfers angular momentum between the inner core and mantle. In a situation where it only experiences this gravitational torque, the mantle is then also entrained in the eastward direction. Conservation of angular momentum in the planetary rotating frame then requires a sustained westward flow (see below) with respect to the mantle in the uppermost part of the outer core[6,22]. Both the gravitational torque and the net torque exerted by the outer core on the inner core vanish when averaged over time (Methods). This leads to long-term co-rotation between the mantle, inner core and lower outer core, with thermochemical wind gradients maintaining the westward drift in the upper outer core. For a given convective forcing, this drift is maximized if direct coupling between the outer core and the mantle is much weaker than indirect coupling via the inner core, as is the case in the coupled Earth dynamo model.

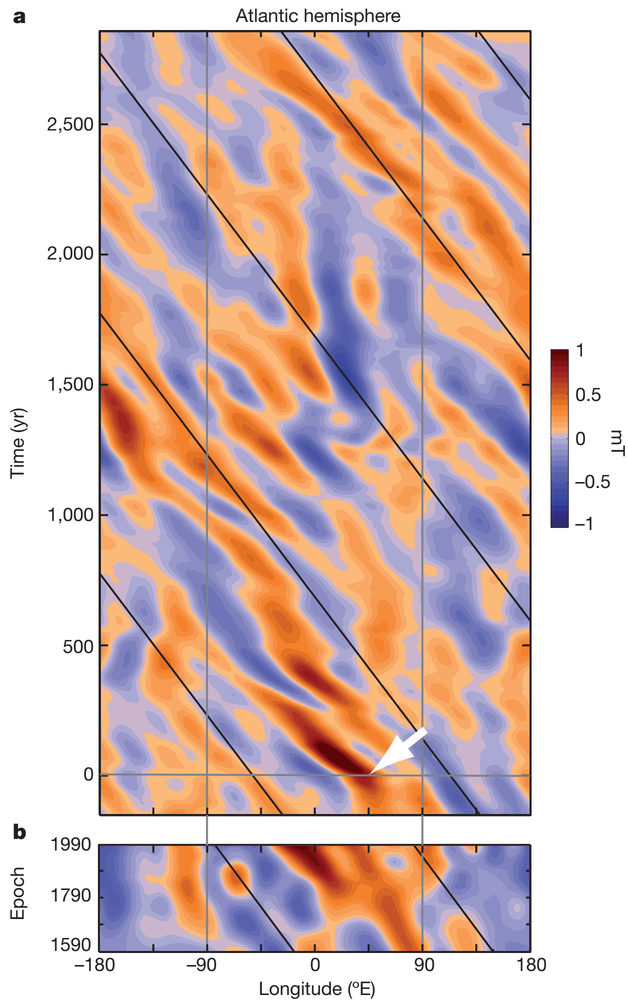


Figure 3 | Temporal evolution of magnetic structures at the Equator. **a**, **b**, Time–longitude plot of the radial magnetic field at the Equator evaluated at the core–mantle boundary (orange denotes an outward magnetic field) from a 3,000-year sequence of the coupled Earth dynamo model filtered to spherical harmonic degree and order 8 (**a**), and from the 400-year sequence of geomagnetic field model[1] gufm1 at its native resolution (**b**). Grey vertical lines mark the Atlantic hemisphere. The grey horizontal line (model time 0) marks the time of the snapshot presented in Figs 2 and 4, and the white arrow marks the reference location. The model time 0 is chosen to be the start of a magnetic patch emergence and drift sequence representative of the localization of secular variation diagnosed in Fig. 1. See Supplementary Video for the evolution of magnetic features at all latitudes in the coupled Earth dynamo model over the entire temporal sequence. The slanted black lines denote a reference westward drift velocity of 10 km yr^{-1}.

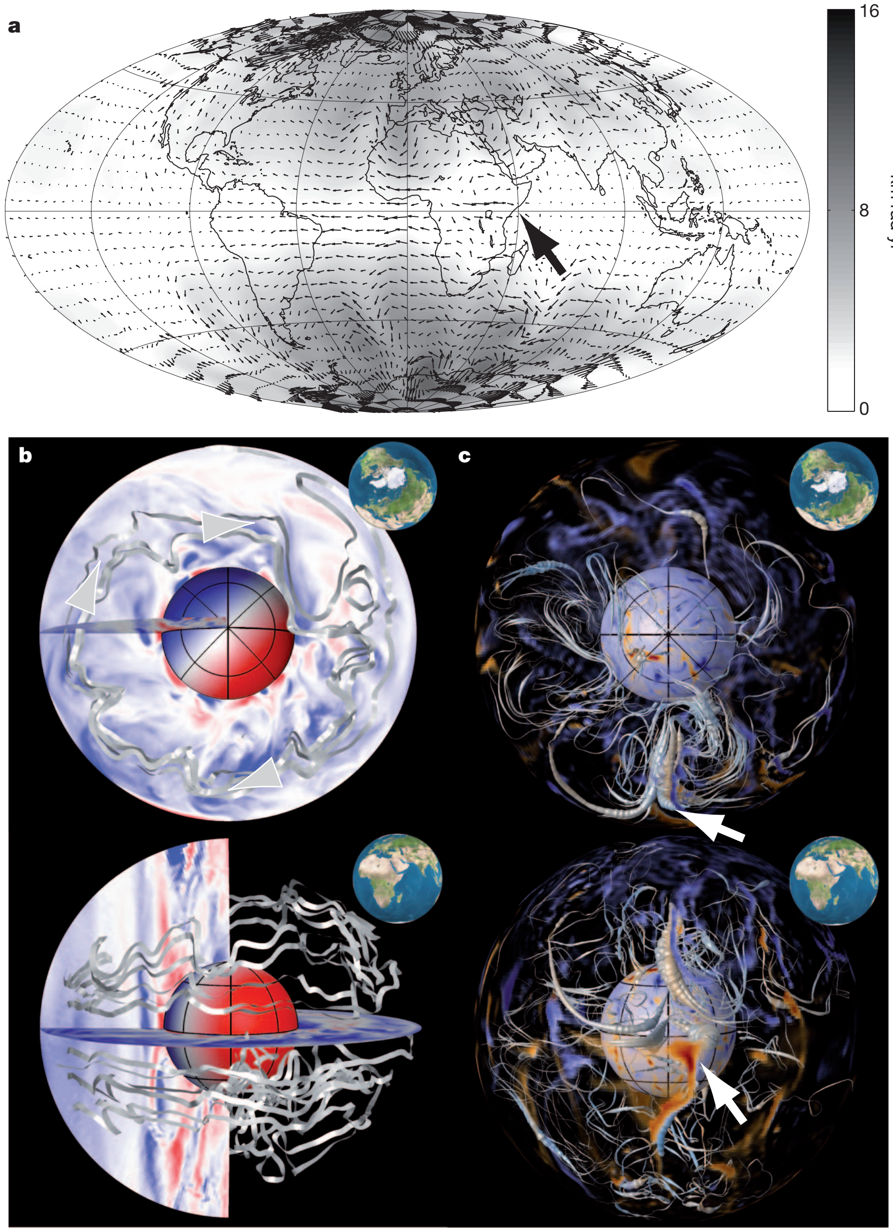


Figure 4 | Internal fluid flow and magnetic structure. **a**, Atlantic-centred Hammer projection of flow at the core–mantle boundary in a snapshot of the coupled Earth dynamo model (small arrows, arbitrary scaling), superimposed on a grey-scale map of the amount of local surface rotation (measured by the absolute toroidal scalar[6], in km rad yr^{-1}). The large black arrow marks the reference location introduced in Figs 2 and 3. **b**, Stream ribbon representations (top, north polar view; bottom, near-equatorial view; insets, indication of the viewing angle) of the internal velocity field structure, superimposed on equatorial and meridional cuts of the azimuthal velocity field (red is eastwards, blue is westwards, maximum 28 km yr^{-1}). Grey arrow heads mark the general flow circulation and the twist of the ribbons is proportional to the local vorticity. The inner-core boundary is colour-coded with the amplitude of the hemispherical modulation in mass anomaly release (red corresponds to excess buoyancy). **c**, Magnetic field line (grey) representations of the internal magnetic structure (same viewing angles as **b**). The field line thickness is proportional to the local magnetic energy density in the shell. Inner and outer boundaries are colour-coded with the amplitude of the unfiltered radial magnetic field (orange is outward, outer boundary is selectively transparent, maxima are 4 mT at the core–mantle boundary and 7 mT at the inner-core boundary). The model snapshot is taken at the same time as in Fig. 2 (which is time 0 in Fig. 3), with white arrows marking the reference location introduced in these figures.

Heterogeneous thermochemical boundary coupling is also included in the coupled Earth dynamo model. We adopt fixed-flux-type[18] conditions for the thermochemical buoyancy both at the core–mantle boundary (where lower-mantle thermal anomalies significantly affect the geodynamo[9,20,25]) and at the inner-core boundary (where solid translation of the inner core[7,8] can spatially modulate the inner-core growth rate, with equally important consequences[6,26,27]). At the core–mantle boundary, a mass anomaly flux pattern[9] derived from lower-mantle seismic tomography[28] is superimposed on a vanishing homogeneous flux. This represents a situation where the upper outer core is neutrally stratified (although our results are robust against a possible stable stratification[29,30], Supplementary Information). At the inner-core boundary, we prescribe a longitudinal hemispheric modulation of mass anomaly release related to inner-core translation. The maximum outward flux is located beneath Indonesia, in resonance with the heterogeneous mantle forcing[9], and as suggested by earlier investigations[6]. Inner-core translation thus proceeds from east to west, opposite to inferences[7] based on upper inner-core seismic texturing. The latter interpretation is however not applicable here according to a parameter space exploration of our model (Supplementary Information and Extended Data Figs 2, 3), which favours

an inner-core translation slower than its mean growth rate. The upper inner-core seismic properties may thus be dominated by solidification texturing[9] rather than texturing induced by the slow ageing of translating iron grains.

By including these elements, the coupled Earth dynamo model is able to produce magnetic variations that are dominated by the emergence of intense, westward-drifting, equatorial flux patches of normal (the same as the ambient) polarity under the Atlantic hemisphere (Supplementary Video). The temporal sequence of emergence, drift and diffusion of these patches agrees between the model and Earth by virtue of the realistic ratio between magnetic induction and diffusion timescales. The associated distribution of average secular variation energy (Fig. 1) quantitatively reproduces the longitudinal and latitudinal localizations of geomagnetic secular variation which standard geodynamo models fail to manage. We find (Extended Data Fig. 4) that latitudinal localization results primarily from the introduction of inner-core/mantle gravitational coupling, whereas longitudinal localization is mainly a consequence of heterogeneous thermochemical coupling between the inner and outer core. Control from the heterogeneous mantle has a secondary effect (Supplementary Information), operating constructively with the inner-core control to promote low secular variation in the Pacific[20]. The instantaneous field and secular variation at the core–mantle boundary exhibit detailed (Fig. 2c, d) and quantitative (Extended Data Fig. 1) agreement with present-day geomagnetic maps derived from high-resolution satellite data (Fig. 2a, b). A time–longitude plot of the evolution of radial magnetic field at the Equator from the coupled Earth dynamo model (Fig. 3a) closely resembles that obtained during the past 400 years (ref. 1; Fig. 3b), with intense patches coherently moving westwards through much of the Atlantic hemisphere, while in the Pacific hemisphere significantly weaker patches undergo slower westward drift with occasional standing and eastward-drifting phases (see also Supplementary Video). The peak of azimuthally moving power (Extended Data Fig. 5) is reached at the Equator, with a coherent westward drift at 14 km yr^{-1}, comparable to the observed speeds[3] of 17 km yr^{-1}. In contrast, the standard dynamo model of Fig. 1 yields a weak and erratic drift at low latitudes with a small amount of power moving slowly eastwards.

The longitudinal localization of magnetic variations in the coupled Earth dynamo model is underlain by a fluid flow (Fig. 4a, b) that agrees remarkably well with recent observation-based inferences of the core flow[6,10,11]. This enables a new, dynamically consistent, interpretation of the physical processes underlying such kinematic flow maps. The gyre driving the westward drift has a sheet-like structure with invariance along the Earth's rotation axis. It produces significant shear at low latitudes close to the core–mantle boundary, concentrating energetic azimuthal magnetic field lines in this region (Fig. 4c). The outer-core upwelling pattern caused by the inner-core hemispherical driving distorts the gyre into an eccentric path that reaches the core–mantle boundary beneath the Atlantic. Expulsion of azimuthal magnetic flux by columnar convection is thus strongest at this location and forms intense radial magnetic flux patches which are subsequently swept westwards by the background flow. The gyre itself is advected westwards, quasi-periodically entering into resonance with the buoyancy heterogeneity of the inner-core boundary. Its shape thus modulates between more eccentric (for example, time 0 in Fig. 3, and Figs 2 and 4) and less eccentric phases with weaker westward drift and magnetic patches more evenly distributed in longitude (for example, time 1,400 in Fig. 3). Assuming that the geodynamo is currently in a phase of more intense gyre eccentricity, our results thus indicate that advection by a planetary-scale thermochemical wind is the origin of low-latitude geomagnetic secular variation. The mechanisms identified in the coupled Earth dynamo model are generic processes related to boundary coupling and involve a competition between the Coriolis, buoyancy and magnetic forces. As realistic relative proportions of these forces are maintained in the simulation over the centennial timescale range relevant for the convective processes (Supplementary Information), these mechanisms should continue to operate in the regime of the Earth's core.

Immediate implications of our findings include the possibility of formulating a data assimilation framework based on a numerical dynamo that includes such boundary couplings. A dynamically consistent prediction of the future geomagnetic field evolution is in sight, and improved knowledge of the past of the core may be accessible through the reanalysis of historical and archaeomagnetic field data[15]. Our results also invite a re-evaluation of the possible strength of inner-core translation with respect to its homogeneous growth. Strong heterogeneous forcing from the inner core is in contradiction with observed geomagnetic secular variation, as it leads to core surface flows and magnetic structures with the reverse drift direction (Extended Data Fig. 3).

METHODS SUMMARY

We solve[31] for Boussinesq convection, thermochemical buoyancy transport and magnetic induction in a system with constant axial rotation at a rate Ω (defining the planetary rotating frame), comprising a spherical fluid shell between radii r_i and r_o ($r_\mathrm{i}/r_\mathrm{o} = 0.35$), a solid inner core of radius r_i with the same electrical conductivity as the fluid shell, and an insulating solid shell representing the mantle between radii r_o and $1.83r_\mathrm{o}$. The inner-core boundary is of a no-slip, electrically conducting type, and axially rotates at a rate Ω_i under the combined influence of viscous, gravitational and magnetic torques[23]. The core–mantle boundary is of a free-slip, electrically insulating type and axially rotates at a rate Ω_m under the influence of the gravitational torque. The gravitational torque on the inner core due to the mantle is given by[22,23] $\Gamma_\mathrm{G} = -\Gamma\tau(\Omega_\mathrm{i} - \Omega_\mathrm{m})$, where $\Gamma\tau = 2.9 \times 10^{20}$ N m yr, Γ being the gravitational coupling constant and τ the viscous relaxation time of the inner core. The inner-core boundary mass anomaly flux is fixed in the frame rotating with the mantle, with a homogeneous contribution f and a longitudinally hemispherical heterogeneity $\Delta f = 0.8f$ maximal[6] at longitude 90° E. The core–mantle boundary mass anomaly flux is fixed in the frame rotating with the mantle, with a vanishing homogeneous contribution and an heterogeneous pattern[9] derived from lower-mantle seismic tomography[28], of peak-to-peak amplitude $\Delta f_\mathrm{o} = 0.115f$. The Ekman number is set to $\mathrm{Ek} = \nu/\Omega D^2 = 3 \times 10^{-5}$ (here ν is the fluid viscosity and $D = r_\mathrm{o} - r_\mathrm{i}$ is the fluid shell depth). The mass anomaly flux Rayleigh number is $\mathrm{Ra_F} = g_\mathrm{o}f/\rho\Omega^3D^2 = 9.3 \times 10^{-5}$ (here g_o is the gravity at the core–mantle boundary and ρ is the fluid density). The ratios between the fluid viscosity, thermal diffusivity κ and magnetic diffusivity λ are set to $\mathrm{Pr} = \nu/\kappa = 1$ and $\mathrm{Pr_m} = \nu/\lambda = 2.5$ (here Pr and $\mathrm{Pr_m}$ are respectively the Prandtl number and the magnetic Prandtl number). The non-dimensional model output is rescaled to the dimensional world using scaling principles thought to hold in both model and Earth: secular variation time scaling[32] and convective-power-based magnetic field scaling[33].

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 March; accepted 14 August 2013.

1. Jackson, A., Jonkers, A. R. T. & Walker, M. R. Four centuries of geomagnetic secular variation from historical records. *Phil. Trans. R. Soc. Lond. A* **358,** 957–990 (2000).
2. Holme, R., Olsen, N. & Bairstow, F. L. Mapping geomagnetic secular variation at the core-mantle boundary. *Geophys. J. Int.* **186,** 521–528 (2011).
3. Finlay, C. C. & Jackson, A. Equatorially dominated magnetic field change at the surface of Earth's core. *Science* **300,** 2084–2086 (2003).
4. Christensen, U. R., Aubert, J. & Hulot, G. Conditions for Earth-like geodynamo models. *Earth Planet. Sci. Lett.* **296,** 487–496 (2010).
5. Buffett, B. Gravitational oscillations in the length of day. *Geophys. Res. Lett.* **23,** 2279–2282 (1996).
6. Aubert, J. Flow throughout the Earth's core inverted from geomagnetic observations and numerical dynamo models. *Geophys. J. Int.* **192,** 537–556 (2013).
7. Monnereau, M., Calvet, M., Margerin, L. & Souriau, A. Lopsided growth of Earth's inner core. *Science* **328,** 1014–1017 (2010).
8. Alboussière, T., Deguen, R. & Melzani, M. Melting-induced stratification above the Earth's inner core due to convective translation. *Nature* **466,** 744–747 (2010).
9. Aubert, J., Amit, H., Hulot, G. & Olson, P. Thermochemical flows couple the Earth's inner core growth to mantle heterogeneity. *Nature* **454,** 758–761 (2008).
10. Pais, M. A. & Jault, D. Quasi-geostrophic flows responsible for the secular variation of the Earth's magnetic field. *Geophys. J. Int.* **173,** 421–443 (2008).
11. Gillet, N., Pais, M. A. & Jault, D. Ensemble inversion of time-dependent core flow models. *Geochem. Geophys. Geosyst.* **10,** Q06004 (2009).
12. Olsen, N., Mandea, M., Sabaka, T. J. & Toffner-Clausen, L. The CHAOS-3 geomagnetic field model and candidates for the 11th generation IGRF. *Earth Planets Space* **62,** 719–727 (2010).
13. Lesur, V., Wardinski, I., Hamoudi, M. & Rother, M. The second generation of the GFZ Reference Internal Magnetic Model: GRIMM-2. *Earth Planets Space* **62,** 765–773 (2010).

14. Finlay, C. C., Jackson, A., Gillet, N. & Olsen, N. Core surface magnetic field evolution 2000–2010. *Geophys. J. Int.* **189,** 761–781 (2012).
15. Fournier, A. *et al.* An introduction to data assimilation and predictability in geomagnetism. *Space Sci. Rev.* **155,** 247–291 (2010).
16. Christensen, U. R. & Wicht, J. in *Treatise on Geophysics* Vol. 8, *Core Dynamics* (ed. Schubert, G.) Ch. 8 (Elsevier, 2007).
17. Kuang, W. & Bloxham, J. An Earth-like numerical dynamo model. *Nature* **389,** 371–374 (1997).
18. Sakuraba, A. & Roberts, P. H. Generation of a strong magnetic field using uniform heat flux at the surface of the core. *Nature Geosci.* **2,** 802–805 (2009).
19. Miyagoshi, T., Kageyama, A. & Sato, T. Zonal flow formation in the Earth's core. *Nature* **463,** 793–796 (2010).
20. Christensen, U. R. & Olson, P. Secular variation in numerical geodynamo models with lateral variations of boundary heat flux. *Phys. Earth Planet. Inter.* **138,** 39–54 (2003).
21. Glatzmaier, G. A. & Roberts, P. H. Rotation and magnetism of Earth's inner core. *Science* **274,** 1887–1891 (1996).
22. Buffett, B. A. & Glatzmaier, G. A. Gravitational braking of inner-core rotation in geodynamo simulations. *Geophys. Res. Lett.* **27,** 3125–3128 (2000).
23. Aubert, J. & Dumberry, M. Steady and fluctuating inner core rotation in numerical geodynamo models. *Geophys. J. Int.* **184,** 162–170 (2011).
24. Dumberry, M. Geodynamic constraints on the steady and time-dependent inner core axial rotation. *Geophys. J. Int.* **170,** 886–895 (2007).
25. Gubbins, D., Sreenivasan, B., Mound, J. & Rost, S. Melting of the Earth's inner core. *Nature* **473,** 361–363 (2011).
26. Davies, C. J., Silva, L. & Mound, J. E. On the influence of a translating inner core in models of outer core convection. *Phys. Earth Planet. Inter.* **214,** 104–114 (2013).
27. Olson, P. & Deguen, R. Eccentricity of the geomagnetic dipole caused by lopsided inner core growth. *Nature Geosci.* **5,** 565–569 (2012).
28. Masters, G., Laske, G., Bolton, H. & Dziewonski, A. in *Earth's Deep Interior* (eds Karato, S., Forte, A., Liebermann, R. C., Masters, G. & Stixrude, L.) 63–87 (AGU Monograph Vol. 117, American Geophysical Union, 2000).
29. Helffrich, G. & Kaneshima, S. Outer-core compositional stratification from observed core wave speed profiles. *Nature* **468,** 807–810 (2010).
30. Gubbins, D. & Davies, C. J. The stratified layer at the core-mantle boundary caused by baro-diffusion of oxygen, sulphur and silicon. *Phys. Earth Planet. Inter.* **215,** 21–28 (2013).
31. Aubert, J., Labrosse, S. & Poitou, C. Modelling the palaeo-evolution of the geodynamo. *Geophys. J. Int.* **179,** 1414–1428 (2009).
32. Lhuillier, F., Fournier, A., Hulot, G. & Aubert, J. The geomagnetic secular-variation timescale in observations and numerical dynamo models. *Geophys. Res. Lett.* **38,** L09306 (2011).
33. Christensen, U. R. & Aubert, J. Scaling properties of convection-driven dynamos in rotating spherical shells and application to planetary magnetic fields. *Geophys. J. Int.* **166,** 97–114 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the French Agence Nationale de la Recherche (grant ANR-2011-BS56-011). Numerical computations were performed at S-CAPAD, IPGP, France, and using HPC resources from GENCI-IDRIS (grants 2012-042122 and at 2013-042122). This is IPGP contribution 3419.

Author Contributions J.A. designed the project and carried out the numerical experiments. J.A. and C.C.F. designed the numerical experiments and processed the results. J.A, C.C.F. and A.F. discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A. (aubert@ipgp.fr).

LETTER

doi:10.1038/nature12525

Key role of symbiotic dinitrogen fixation in tropical forest secondary succession

Sarah A. Batterman[1], Lars O. Hedin[1], Michiel van Breugel[2], Johannes Ransijn[3,4], Dylan J. Craven[5]† & Jefferson S. Hall[2]

Forests contribute a significant portion of the land carbon sink, but their ability to sequester CO_2 may be constrained by nitrogen[1–6], a major plant-limiting nutrient. Many tropical forests possess tree species capable of fixing atmospheric dinitrogen (N_2)[7], but it is unclear whether this functional group can supply the nitrogen needed as forests recover from disturbance or previous land use[1], or expand in response to rising CO_2 (refs 6, 8). Here we identify a powerful feedback mechanism in which N_2 fixation can overcome ecosystem-scale deficiencies in nitrogen that emerge during periods of rapid biomass accumulation in tropical forests. Over a 300-year chronosequence in Panama, N_2-fixing tree species accumulated carbon up to nine times faster per individual than their non-fixing neighbours (greatest difference in youngest forests), and showed species-specific differences in the amount and timing of fixation. As a result of fast growth and high fixation, fixers provided a large fraction of the nitrogen needed to support net forest growth (50,000 kg carbon per hectare) in the first 12 years. A key element of ecosystem functional diversity was ensured by the presence of different N_2-fixing tree species across the entire forest age sequence. These findings show that symbiotic N_2 fixation can have a central role in nitrogen cycling during tropical forest stand development, with potentially important implications for the ability of tropical forests to sequester CO_2.

The recognition that large proportions of tropical forests are recovering from cutting, agricultural land use or natural disturbances[1,9,10] raises fundamental questions about limits on the tropical carbon sink. Nitrogen may constrain the post-disturbance recovery of vegetation and soil carbon pools[1,3,11,12] and may limit forest response[6,8] to rising CO_2, but the extent to which biological N_2 fixation can relieve these constraints remains unresolved[3,13,14]. Tree species capable of symbiotic N_2 fixation are abundant in many tropical forests, but recent observations show that, depending on local conditions, individual trees may or may not produce the symbiotic nodules where fixation occurs[15], that phosphorus may ultimately constrain fixation in tropical soils[8,16–18] and that some tropical forests contain a low diversity of N_2-fixing species[7,19].

It is therefore crucial to resolve whether N_2-fixing trees can relieve nitrogen limitation and ensure fixation as an ecosystem service[20] during tropical forest secondary succession. The existence of a feedback between post-disturbance forest regrowth and N_2 fixation could influence not only the limits to forest CO_2 uptake but also how we understand and model potentially nonlinear interactions between land use, nitrogen cycling and the land carbon sink[2,3,21].

Here we use a replicated, natural ecosystem experiment of forest recovery from land use across Panamanian forests to ask, first, whether symbiotic N_2 fixation can overcome the nitrogen deficiency imposed by rapid biomass accumulation during secondary succession; and, second, whether ecosystem-scale fixation depends on the identity and diversity of N_2-fixing species.

The Agua Salud Project encompasses approximately 15 km^2 of tropical moist forests[22] on phosphorus-poor oxisols similar to soils in large tropical regions worldwide[23] (Supplementary Note 1 and Supplementary Table 1). We examined 12 randomly chosen 0.2-ha plots abandoned from pasture 5, 12 or 30 years ago (four replicate plots per age), plots abandoned approximately 80 years ago (two 1.0-ha plots), and forests near mature status (approximately 300 years; two 1.0-ha plots).

We measured diameters of 13,372 trees from 297 identified tree species, and collected 244 trees to develop allometric biomass equations for the 26 dominant species[22] (Methods, Supplementary Note 2 and Supplementary Table 2). Fixers constituted 2.6–14.5% of basal area (Table 1), which is at the low end for many tropical forests, and several of our most abundant species are widespread throughout the neotropics[7]. Comparison between 2009 and 2011 measures enabled us to quantify directly individual tree growth rates and changes in community composition. Biomass carbon pools were calculated from tree allometries (sum of leaves, wood and roots; Supplementary Note 2, Supplementary Fig. 1 and Supplementary Table 2); leaf nutrients across 55 species were similar to other forests[24].

We determined fixation in 268 trees (approximately 30% of fixers in all plots) across the nine most abundant species (>70% of fixer basal area across plots; Supplementary Notes 3 and 4). For each tree we quantified active N_2-fixing nodules beneath the canopy (nodulation; $g\,m^{-2}$) and estimated the fixation rate based on ^{15}N-tracer incubations or acetylene reduction (Supplementary Note 3 and Supplementary Table 4). This enabled us to scale up fixation to the ecosystem, by summing across all individual N_2 fixers in all plots using species- and forest-age-specific data on nodulation and per-nodule biomass fixation (scenario number 4 in Supplementary Note 3).

We observed rapid forest regrowth and biomass carbon accumulation across the chronosequence (Fig. 1a and Supplementary Table 3). The rate of biomass accumulation (that is, net biomass carbon increase per area and forest age) was exceptionally high in young forests, such that our 12-year forests contained >40% (>50,000 $kg\,C\,ha^{-1}$; >4,100 $kg\,C\,ha^{-1}\,yr^{-1}$) of the biomass present in the 300-year forests (approximately 120,500 $kg\,C\,ha^{-1}$). Biomass carbon accumulation decelerated in older forests (black line, Fig. 1a), with a theoretical long-term equilibrium biomass that compared well with eleven old-growth Panamanian forests[25] (red line, Fig. 1a).

We also observed a dramatic trend in N_2 fixation across the forest chronosequence (Fig. 1b, Supplementary Fig. 2 and Supplementary Table 5). Ecosystem nodulation and total N_2 fixation increased steeply to a maximum in our 12-year forests (29 $kg\,N\,ha^{-1}\,yr^{-1}$), and then decreased to low levels in 80-year forests. At 300 years, fixation recovered to 8% of maximum (Table 1, using scenario number 4 in Supplementary Note 3). Because our sampling was limited to specific forest ages, we cannot evaluate whether even higher fixation may have occurred between 5 and 30 years of forest age.

Highest nodulation and fixation coincided with the period of most rapid plant biomass accumulation (Fig. 1a), implying that nodulation and N_2 fixation were induced by high demand for nitrogen imposed by

[1]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA. [2]Smithsonian Tropical Research Institute, Balboa, Ancón, Panamá, Panama. [3]Forest Ecology and Forest Management Group, Centre for Ecosystem Studies, Wageningen University, 6700 AA Wageningen, The Netherlands. [4]Department of Geosciences and Natural Resource Management, University of Copenhagen, Rolighedsvej 23, DK-1958 Frederiksberg C, Denmark. [5]School of Forestry and Environmental Studies, Yale University, New Haven, Connecticut 06511, USA. †Present address: German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany.

Table 1 | Properties of N_2 fixers and fixation at the ecosystem, community and individual scales during succession

Variable	Forest age				
	5 years (4 plots)	12 years (4 plots)	30 years (4 plots)	80 years (2 plots)	300 years (2 plots)
Ecosystem: total N_2 fixation (kg N ha^{-1} yr^{-1})	10 (4)	29 (12)	11 (6)	0.3 (0.1)	2.2 (0.2)
Community: fixer basal area (per cent of total)	5.0 (1.7)	14.5 (1.1)	5.4 (2.3)	2.6 (0.9)	5.8 (1.4)
Community: fixer versus non-fixer recruitment (per cent per 2 years)	61 versus 5.2	0.51 versus −2.9	−1.9 versus −1.15	NA	NA
Individual: ratio of fixer to non-fixer growth rate	9.2 (3.7)	2.3 (0.6)	1.1 (0.9)	NA	NA

Numbers represent mean (s.e.m.). Ecosystem total N_2 fixation is calculated based on species- and forest-age-specific data for both nodulation and per-nodule biomass fixation (scenario number 4 in Supplementary Note 3). Recruitment was calculated as the per cent of trees remaining in 2011 compared to 2009. Positive numbers indicate increasing abundance (that is, positive recruitment), and negative numbers signify that the group is decreasing (that is, self-thinning). NA, numbers not available because the forest inventory was limited to 1 year. Growth rate was calculated as the average difference in per-individual biomass between 2009 and 2011. Values of greater than 1 indicate greater growth of fixers relative to non-fixers, whereas values of less than 1 indicate that fixers are growing slower relative to non-fixers.

the regrowing forest. In turn, fixation decreased between the 12-, 30- and 80-year-old forests, coincident with a progressive slowdown of biomass accumulation; fixation in 300-year-old forests was similar to independent measures[15] from nearby old-growth forests (red point, Fig. 1b).

These ecosystem trends in N_2 fixation were corroborated by independent observations at scales of plant community composition and individual plant growth (Table 1). At the community scale, relative fixer abundance (per cent of total basal area) displayed the same trend as ecosystem nodulation and fixation: a steep increase to a peak in 12-year forests, followed by a decrease to low levels in forests of 80 years and older. At the individual scale, fixers accumulated biomass ninefold faster than non-fixers at 5 years into forest recovery, but the fixer growth advantage declined to twofold in 12-year-old forests, and became equivalent to non-fixers in 30-year-old forests. This fixer growth advantage was also apparent at the community scale in that recruitment of fixers greatly exceeded non-fixers (61% versus 5%) in young forests, but decreased to similar levels in the progressively older forests. The partial recovery of fixation and percentage basal area of fixers in 300-year-old forests was due to increased abundance and per-individual fixation of *Tachigali versicolor* (discussed below; see also Supplementary Note 4).

We next asked how these emergent patterns in fixation depended upon the identity of N_2-fixing species. When assessed across all plots, individual species ranged approximately 200-fold in total nodulation and fixation (Fig. 2a and Table 2), with two species contributing >65% of all nodulation: *Inga cocleensis* and *T. versicolor*.

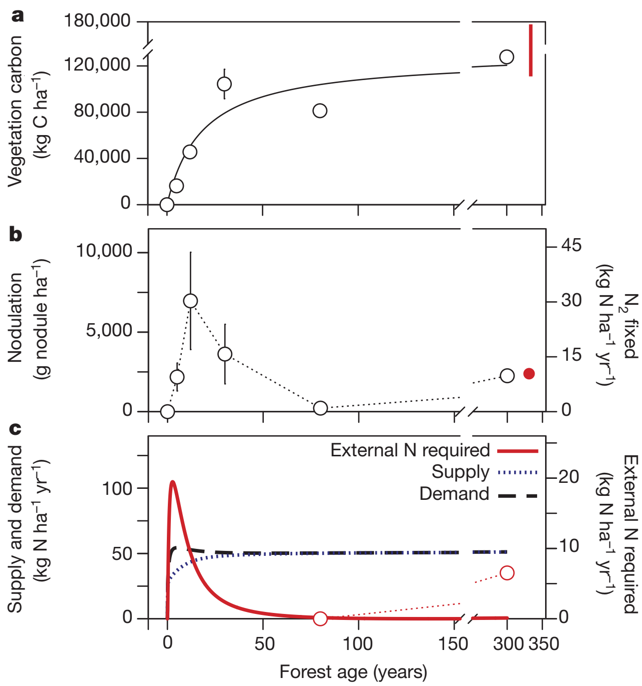


Figure 1 | Biomass carbon accumulation and dinitrogen fixation in forests recovering from land use. **a**, Sum of above- and belowground vegetation carbon (circles). Error bars (means ± s.e.m.) are invisible when replicates were similar. The fitted Michaelis–Menten curve was used to determine net nitrogen demand for the model, below. Red bar identifies the biomass range observed across 11 (ref. 25) old-growth Panamanian forests. **b**, Nodule biomass and N_2 fixed (mean ± s.e.m.). Fixation differs slightly from Table 1 because constant per-nodule biomass fixation rate is assumed (scenario number 1, Supplementary Note 3). Red point identifies fixation in nearby mature forests[15]. **c**, Modelled plant nitrogen demand (black), soil nitrogen supply (blue) and external nitrogen required from N_2 fixation (red). Dotted red line illustrates that fixation increases with disturbance in mature forests. $n = 4$ 0.2-ha plots for 5–30 years, $n = 2$ 1.0-ha plots for 80–300 years.

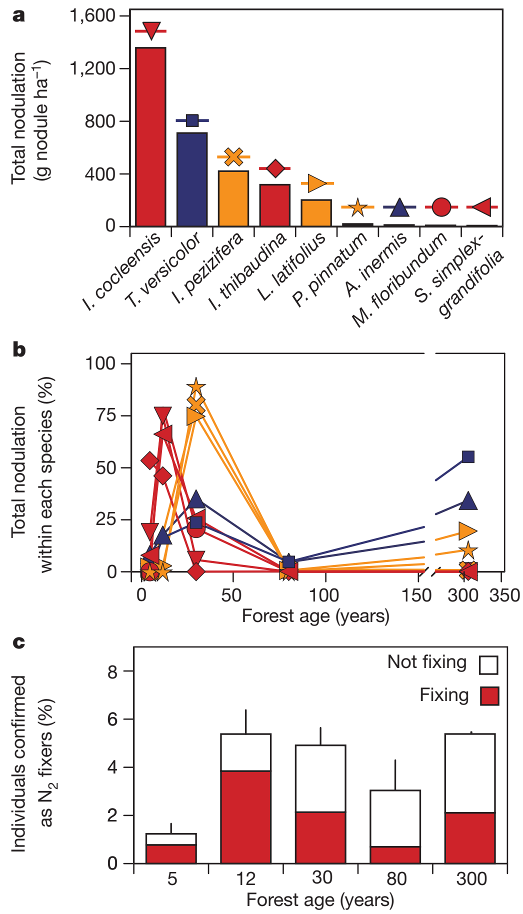


Figure 2 | Individual species contributions to ecosystem fixation. **a**, Calculated total nodulation by species: *Inga cocleensis* ($n = 64$), *Tachigali versicolor* ($n = 36$), *Inga pezizifera* ($n = 29$), *Inga thibaudiana* ($n = 30$), *Lonchocarpus latifolius* ($n = 21$), *Platymiscium pinnatum* ($n = 12$), *Andira inermis* ($n = 30$), *Machaerium floribundum* ($n = 11$) and *Swartzia simplex-grandifolia* ($n = 35$). **b**, Observed percentage of total fixation for each species at each forest age, and the species that specialize on young (red), mid-successional (orange) and mature (blue) forests (colours and symbols are same as in **a**). n values are given in Supplementary Data. **c**, Proportion of trees with known ability to fix that are currently fixing (red) and not fixing (white). $n = 4$ 0.2-ha plots for 5–30 years, $n = 2$ 1-ha plots for 80–300 years. Error bars indicate s.e.m.

Although six of the nine N_2-fixing species were present throughout the chronosequence, individual species differed in when they fixed the most and were most abundant (Fig. 2b and Table 2). One group dominated by *I. cocleensis* fixed primarily in the youngest forests, a second group in 30-year-old forests (dominated by *Inga pezizifera*), and a third group in mature forests (dominated by *T. versicolor*). We infer that these forests depend on a diverse suite of tree species that partition the biogeochemical niche[26] for fixation over the course of biomass recovery.

We further asked whether our observed fixation trend was generated by, first, N_2-fixing species that always fix but are progressively excluded from the community as nitrogen is replenished (that is, obligate fixation); or, second, N_2-fixing species that persist over time but that turn fixation on or off depending on their nitrogen balance (that is, facultative fixation). This distinction is significant because the first theory implies that the ecosystem capacity for fixation disappears over time, whereas the second theory implies that the capacity is functionally ensured throughout forest succession.

Our results support the second theory. The proportion of trees with confirmed ability to fix remained relatively stable in forests 12 years and older, yet the proportion that expressed fixation declined from 71% at 12 years to 23% at 80 years (Fig. 2c and Table 2). Furthermore, even when fixing, individuals varied their investment in fixation and nodulation across at least two orders of magnitude independent of species (Table 2), consistent with facultative fixation at the individual scale. These dynamics indicate that fixers can buffer forest nitrogen supply by upregulating fixation when nitrogen is low[15,27].

Our findings identify an ecosystem-scale feedback between forest regrowth, vegetation nitrogen demand and the supply of new nitrogen ensured by N_2 fixers. The results imply that nitrogen limitation is the mechanism favouring fixers and fixation in young forests, but the drop in fixation in older forests could result from either an alleviation of nitrogen-limitation or a decrease in light available to fixers. However, light does not explain the approximately 99% decline between maximum and minimum fixation rates (12 versus 80 years), because the proportion of fixers with canopy access to sunlight remained high over time (51% versus 45%, respectively; Supplementary Note 8 and Supplementary Table 7).

We evaluated the nitrogen feedback using a mass-balance model that captures the mechanisms of plant nitrogen demand (net leaf, wood and fine-root growth), plant–soil nitrogen recycling, nitrogen inputs (deposition and fixation) and nitrogen losses (Methods, Supplementary Notes 5–7, Supplementary Box 1, Supplementary Table 6 and Supplementary Figs 3, 6 and 7). When the model was run with the observed pattern of biomass carbon and nitrogen accumulation (fitted line in Fig. 1a), it revealed, first, that the nitrogen demand imposed by vegetation growth greatly exceeded the capacity of internal recycling to supply plant nitrogen in the first approximately 25 years (dashed versus dotted lines, Fig. 1c); second, that the resulting nitrogen deficit (red line, Fig. 1c) was highest when biomass accumulation was steepest in early succession, and declined as biomass accumulation decelerated over time (red line in Fig. 1c versus black line in Fig. 1a); and, third, that the overall shape of the resulting nitrogen deficit recreated the observed changes in fixation across our plots (red line in Fig. 1c versus black line in Fig. 1b). The field data indicated a lag in fixation in our youngest forests, compared to the model, probably caused by slow establishment and growth of the early tree community (Supplementary Note 6 and Supplementary Figs 4 and 5).

Although fixation would be downregulated as the plant–soil nitrogen cycle becomes increasingly balanced (Fig. 1c), we nevertheless observed low but continued fixation in 300-year-old forests by the gap specialist, *T. versicolor*. This indicates that the feedback between

Table 2 | Properties of individual species at different stages of forest recovery

Forest age (years)	Species (from left to right in decreasing order of importance for N_2 fixation at the ecosystem level)								
	I. cocleensis	*T. versicolor*	*I. pezizifera*	*I. thibaudiana*	*L. latifolius*	*P. pinnatum*	*A. inermis*	*M. floribundum*	*S. simplex-grandifolia*
Abundance, given as the number of trees per species per ha; the total number of trees of any species per ha is given in parentheses beside forest age									
5 (13,806)	230	0	10	135	65	0	95	0	30
12 (6,145)	460	5	80	150	235	0	25	7	140
30 (3,925)	70	95	100	15	5	60	70	8	165
80 (2,240)	8	31	22	7	5	3	4	0	5
300 (1,569)	9	48	12	6	18	6	24	0	0
Trees that were N_2-fixing (per cent of trees examined)									
5	62	NA	NA	82	83	NA	17	NA	100
12	91	100	81	87	60	NA	50	22	40
30	56	53	54	100	100	72	17	50	25
80	0	48	33	NA	25	0	17	NA	0
300	8	72	25	NA	45	50	30	NA	0
Mean nodulation rate for N_2-fixing trees (min–max) (g nodule m^{-2} under tree canopy)									
5	1.02 (0.06–9.34)	NA	NA	1.07 (0.11–6.66)	1.17 (0.04–3.00)	NA	0.18	NA	0.82
12	1.93 (0.15–7.79)	8.12	1.35 (0.2–12.87)	1.43 (0.30–8.40)	0.34 (0.18–0.49)	NA	0.15	1.54 (0.15–16.01)	0.63 (0.08–3.66)
30	1.81 (0.02–4.78)	6.73 (0.30–31.87)	3.09 (0.25–8.03)	1.18 (0.3–3.78)	9.05	1.08 (0.27–2.26)	0.27 (0.21–0.36)	4.94	0.55 (0.11–3.23)
80	NA	1.82 (0.44–2.95)	0.79	NA	0.19	NA	0.11	NA	NA
300	3.33	3.06 (0.18–16.01)	0.46	NA	0.99 (0.32–5.62)	1.18 (0.14–9.96)	0.18 (0.06–0.39)	NA	NA
Mean fixation rate for N_2-fixing trees* (s.e.m.) (g N m^{-2} yr^{-1} under tree canopy)									
5	4.6 (2.2)	NA	NA	4.8 (2.1)	0.8 (0.8)	NA	0.5 (NA)	NA	2.2 (NA)
12	8.7 (2.4)	7.4 (NA)	6.5 (2.8)	6.5 (2.8)	0.2 (0.1)	NA	0.4 (NA)	4.1 (NA)	1.7 (0.3)
30	8.2 (4.8)	6.1 (3.6)	5.3 (NA)	5.3 (NA)	6.4 (NA)	2.9 (1.4)	0.7 (NA)	13.2 (NA)	1.5 (1.2)
80	NA	1.7 (0.3)	2.6 (NA)	NA	0.1 (NA)	NA	0.3 (NA)	NA	NA
300	11.1 (NA)	2.8 (1.1)	1.6 (NA)	NA	0.7 (0.2)	3.1 (NA)	0.5 (NA)	NA	NA

* Uses the species- and forest-age-specific per-nodule biomass fixation rate. NA, data not available because the species was either absent or not measured, or because the error could not be calculated due to insufficient replication. The number of trees sampled for Table 2 is given in the Supplementary Data for Fig. 2b.

fixation and soil nitrogen deficiency also can occur in tree-fall gaps[15] that begin to form as stands mature and become less uniform; *T. versicolor* grows rapidly when such light gaps form.

Our findings offer an example in which a vital ecosystem function emerges from the properties of individual species. We estimate that fixers provided >50% of the nitrogen needed to support the sequestration of 50,000 kg of plant carbon per hectare over the first 12 years of forest recovery. The presence of such high fixation rates on the phosphorus-poor Agua Salud soils implies that phosphorus may be less of a constraint on tropical fixation than often assumed in theories or models. Other factors including competition for light or soil molybdenum[28] may ultimately limit fixers and/or fixation. Most fundamentally, our results offer a mechanism by which tropical forests can avoid nitrogen constraints caused by disturbance or land-use transition, or that are predicted to occur in response to rising atmospheric CO_2.

METHODS SUMMARY

We sampled 16 forest plots at different stages of recovery from agriculture to intact forest within the 15 km^2 Agua Salud Project and in the nearby Soberania National Park. All plots had similar climate and were on soils (oxisols) with similar chemical and physical characteristics. In each plot, we measured the diameter of trees and calculated aboveground and belowground carbon and nitrogen biomass using allometric equations developed from 244 trees collected at our sites. We analysed leaf carbon and nitrogen contents from 55 common species across our plots, and used standard conversion factors to estimate carbon and nitrogen biomass across leaf, roots and wood tissues of each individual tree. To determine whole-ecosystem biomass carbon and nitrogen, we summed across all individuals of each species in each plot.

We quantified fixation in individuals and at the ecosystem scale by measuring nodule biomass for 268 trees across the 9 most abundant N_2-fixing species, and per-nodule biomass fixation rates derived from ^{15}N incorporation and acetylene reduction. We calculated fixation at the scale of the ecosystem across both individual species (sum of all individuals of a given species) and across all species combined (sum of all individuals from all N_2-fixing species).

We developed a mass-balance plant–soil ecosystem model in which plant nitrogen demand and plant litter input to soils was externally forced by the observed carbon accumulation curve across our chronosequence. We considered a plant-available and a plant-unavailable soil pool, and a 'passive' soil pool that turned over on a timescale greater than the successional dynamics of this study (and thus was considered external to the plant–soil model). Nitrogen can enter the system through inputs from deposition, the slow mineralization of the externalized passive pool or N_2 fixation, which is determined as the difference between plant nitrogen demand and soil nitrogen supply.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 November 2012; accepted 2 August 2013.
Published online 15 September 2013.

1. Davidson, E. A. *et al.* Recuperation of nitrogen cycling in Amazonian forests following agricultural abandonment. *Nature* **447,** 995–998 (2007).
2. Zaehle, S. & Dalmonech, D. Carbon–nitrogen interactions on land at global scales: Current understanding in modeling climate biosphere feedbacks. *Curr. Opin. Environ. Sustain.* **3,** 311–320 (2011).
3. Gerber, S., Hedin, L. O., Oppenheimer, M., Pacala, S. W. & Shevliakova, E. Nitrogen cycling and feedbacks in a global dynamic land model. *Glob. Biogeochem. Cycles* **24,** GB1001 (2010).
4. Thornton, P. E., Lamarque, J.-F., Rosenbloom, N. A. & Mahowald, N. M. Influence of carbon–nitrogen cycle coupling on land model response to CO_2 fertilization and climate variability. *Glob. Biogeochem. Cycles* **21,** GB4108 (2007).
5. Houghton, R. A., Hall, F. & Goetz, S. J. Importance of biomass in the global carbon cycle. *J. Geophys. Res.* **114,** G00E03 (2009).
6. Hungate, B. A., Dukes, J. S., Shaw, R., Luo, Y. & Field, C. Nitrogen and climate change. *Science* **302,** 1512–1513 (2003).
7. Sprent, J. I. *Legume Nodulation: A Global Perspective*. (Wiley-Blackwell, 2009).
8. Goll, D. S. *et al.* Nutrient limitation reduces land carbon uptake in simulations with a model of combined carbon, nitrogen and phosphorus cycling. *Biogeosci. Discuss.* **9,** 3173–3232 (2012).
9. Brown, S. & Lugo, A. E. Tropical secondary forests. *J. Trop. Ecol.* **6,** 1–32 (1990).
10. Pan, Y. *et al.* A large and persistent carbon sink in the world's forests. *Science* **333,** 988–993 (2011).
11. Davidson, E. A. *et al.* Nitrogen and phosphorus limitation of biomass growth in a tropical secondary forest. *Ecol. Appl.* **14,** S150–S163 (2004).
12. Amazonas, N. T., Martinelli, L. A., Piccolo, M. C. & Rodrigues, R. R. Nitrogen dynamics during ecosystem development in tropical forest restoration. *For. Ecol. Manage.* **262,** 1551–1557 (2011).
13. Russell, A. E. & Raich, J. W. Rapidly growing tropical trees mobilize remarkable amounts of nitrogen, in ways that differ surprisingly among species. *Proc. Natl Acad. Sci. USA* **109,** 10398–10402 (2012).
14. Gehring, C., Vlek, P. L. G., de Souza, L. A. G. & Denich, M. Biological nitrogen fixation in secondary regrowth and mature rainforest of central Amazonia. *Agric. Ecosyst. Environ.* **111,** 237–252 (2005).
15. Barron, A. R., Purves, D. W. & Hedin, L. O. Facultative nitrogen fixation by canopy legumes in a lowland tropical forest. *Oecologia* **165,** 511–520 (2011).
16. Binkley, D., Senock, R. & Cromack, K. Jr. Phosphorus limitation on nitrogen fixation by *Facaltaria* seedlings. *For. Ecol. Manage.* **186,** 171–176 (2003).
17. Vitousek, P. M. & Howarth, R. W. Nitrogen limitation on land and in the sea: how can it occur? *Biogeochemistry* **13,** 87–115 (1991).
18. van Groenigen, K. J. *et al.* Element interactions limit soil carbon storage. *Proc. Natl Acad. Sci. USA* **103,** 6571–6574 (2006).
19. Hedin, L. O., Brookshire, E. N. J., Menge, D. N. L. & Barron, A. R. The nitrogen paradox in tropical forest ecosystems. *Annu. Rev. Ecol. Evol. Syst.* **40,** 613–635 (2009).
20. Isbell, F. *et al.* High plant diversity is needed to maintain ecosystem services. *Nature* **477,** 199–202 (2011).
21. Yang, X., Richardson, T. K. & Jain, A. K. Contributions of secondary forest and nitrogen dynamics to terrestrial carbon uptake. *Biogeosciences* **7,** 3041–3050 (2010).
22. van Breugel, M., Ransijn, J., Craven, D., Bongers, F. & Hall, J. S. Estimating carbon stock in secondary forests: decisions and uncertainties associated with allometric biomass models. *For. Ecol. Manage.* **262,** 1648–1657 (2011).
23. Quesada, C. A. *et al.* Variations in chemical and physical properties of Amazon forest soils in relation to their genesis. *Biogeosciences* **7,** 1515–1541 (2010).
24. Fyllas, N. M. *et al.* Basin-wide variations in foliar properties of Amazonian forest: phylogeny, soils and climate. *Biogeosci. Discuss.* **6,** 3707–3769 (2009).
25. Chave, J. *et al.* Error propagation and scaling for tropical forest biomass estimates. *Phil. Trans. R. Soc. Lond. B* **359,** 409–420 (2004).
26. Menge, D. N. & Hedin, L. O. Nitrogen fixation in different biogeochemical niches along a 120000-year chronosequence in New Zealand. *Ecology* **90,** 2190–2201 (2009).
27. Batterman, S. A., Wurzburger, N. & Hedin, L. O. Nitrogen and phosphorus interact to control tropical symbiotic N_2 fixation: a test in *Inga punctata*. *J. Ecology* (in the press).
28. Barron, A. R. *et al.* Molybdenum limitation of asymbiotic nitrogen fixation in tropical forest soils. *Nature Geosci.* **2,** 42–45 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Adelberg and K. Zelazny for assisting with data collection, M. Baillon and A. Hernandez for botanical identifications, N. Wurzburger and A. Barron for species-specific N_2 fixation rates, J. Sprent for advice about N_2-fixing trees and P. Reich for comments. This work was supported by grants to L.O.H. from the National Science Foundation (NSF; DEB-0614116), the National Oceanic and Atmospheric Association (NOAA; grant NA17RJ262 – 344), the Cooperative Institute for Climate Science of Princeton University and the Carbon Mitigation Initiative of Princeton University; and to S.A.B. from the Smithsonian Tropical Research Institute (STRI). It is a contribution to the Agua Salud Project (ASP), a collaboration among STRI, the Panama Canal Authority (ACP) and the National Environmental Authority of Panama (ANAM). ASP funding came from the HSBC climate partnership, STRI, the Frank Levinson Family Foundation, the Motta Family Foundation and an anonymous donor.

Author Contributions S.A.B., L.O.H., J.S.H. and M.v.B. designed the project. S.A.B. conducted N_2-fixation-related field work. J.R., M.v.B. and J.S.H. provided allometry data; D.J.C. provided plant nutrient data. S.A.B. and L.O.H. wrote the paper. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.A.B. (sbatterm@princeton.edu).

METHODS

Study sites. We sampled 16 forest plots at different stages of recovery from agriculture to intact forest within the 15-km^2 Agua Salud Project[22] and in Soberania National Park approximately 3 km from Agua Salud. Soils across all sites were highly phosphorus-poor and clay-rich oxisols derived from basalt parent material, with no significant differences in carbon, nitrogen, phosphorus or cations across forest age ($P > 0.05$, ANOVA, d.f. = 4 for total carbon, total nitrogen, carbon:nitrogen, total phosphorus; and d.f. = 2 for Mg, K, Ca, Al and extractable phosphorus) (Supplementary Table 1). All sites were classified as tropical moist forest, received similar annual rainfall (2,700 mm; S. Paton, personal communication) with a December-to-April dry season, and shared a diverse community of tree species (>297). Forests contained 24 N_2-fixing tree taxa, many of which are commonly found throughout neotropical forests. We measured all biological and soil properties in the wet season.

Sampling approach. We selected sites to create a replicated sequence of forest recovery ages, with all sites <80 years old recovering from grazing, and sites 80 and 300 years old recovering from either cutting alone or cutting followed by grazing and/or non-intensive agriculture[29,30]. We used a standard chronosequence approach[31] in which space-for-time substitution is used to infer changes in ecosystem properties at temporal scales that exceed the limit of direct observation, and which assumes that sites share similar initial conditions and similar dynamics over time. Our forests and soils are very similar in composition and historical origin, but, depending on forest response to CO_2 fertilization and/or temperature change (at present unknown), increased atmospheric CO_2 or temperature may have introduced a historic bias in the rate of recovery in younger forest stands.

We adjusted area and number of plots to account for differences in tree density (13,800 versus 1,500 trees per ha in 5- versus 300-year-old forests, respectively; Table 2), landscape variability in community composition (high in young forests), biomass accumulation (high in young forests; Fig. 1a) and the influence of localized forest gap dynamics (absent in young forests but common in 300-year-old forests). Note that in mature forest it is not possible to conceptually or visually separate 'gaps' (unless they are only a few years old) from 'mature forest'. We sampled four replicate 0.2-ha plots in all forests <80 years, and two replicate 1.0-ha plots in the 80- and 300-year-old forests. In plots <80 years, we measured every tree with a diameter at breast height (DBH) of ≥5 cm, and half of all trees with a DBH of 1 to 5 cm; in 80- and 300-year-old plots we measured every tree with a DBH of ≥10 cm and two-fifths of trees with a DBH of 1 to 10 cm in 2008. We sampled all plots of <80 years in age in 2009 and 2011; this enabled us to reliably calculate differences in individual tree growth rates between fixers and non-fixers, and to quantify changes in community composition during the period of highest biomass accumulation and strongest competition for light. We did not re-sample our older sites because growth calculated from DBH becomes reliable only in large trees over longer timescales.

Biomass calculations. We used a standard allometric approach to quantify biomass allocation to leaves, wood and fine roots across all plots (Supplementary Fig. 1 and Supplementary Tables 2 and 3). For each of our 26 most common species (contributing approximately 75% of total basal area across all plots) we developed equations (all $R^2 > 0.85$) for aboveground wood and leaf biomass by collecting 244 trees <26 cm DBH: $B_{x,s} = Y_{x,s} \times \mathrm{DBH}^{(Z_{x,s})}$, where B is biomass, Y is a proportionality coefficient, Z is a scaling exponent, x is leaf or wood, and s is species[22] (Supplementary Table 2). For all other species, we used a general equation based on all species combined (Supplementary Table 2). Roots and the sum of aboveground and belowground wood were calculated as described below. For trees with a DBH of ≥26 cm, we used a general equation for aboveground tropical forest biomass[25] (AGB): $\mathrm{AGB} = A \times e^B$, where $B = -3.742 + 3.45 \times \log(\mathrm{DBH}) - 0.148 \times \log(\mathrm{DBH})^2$ and A is a coefficient for wood density. We calculated biomass allocation to leaves versus roots versus wood, as follows: leaf area index equilibrating at approximately 0.2 kg N m^{-2} after canopy closure; fine roots equal to leaf biomass[9,32]; coarse roots from a published allometric relation with stem biomass derived from the global Luyssaert database[33]; and total wood = AGB − leaf biomass + coarse root biomass. Biomass was summed across all trees in each plot and converted to carbon based on the average carbon content (0.47) in our forests[34].

Tissue nutrient pools. We determined biomass, and biomass carbon and nitrogen pools (Supplementary Table 3), by summing the biomass of individual trees (estimated with allometric equations as described above) across all individuals and species and applying the nitrogen content of tissues that we determined from our forests and the literature. We measured species-specific leaf nitrogen content of leaves from the 55 most common species of trees across the Agua Salud forest sites[35]. Leaf nitrogen contents were within the range found at other neotropical sites[24,36]. To calculate the total leaf nitrogen pool size, we found the community abundance-weighted average leaf nitrogen content and applied that to our total leaf biomass pool size. We used a standard nitrogen content of wood derived from the literature[37] in the lower (and conservative, for our calculations) range of tropical wood nitrogen contents[38], and we approximated fine-root nitrogen content as being equal to leaves. Total forest biomass nitrogen was determined as the sum of leaf, wood, and root nitrogen.

Dinitrogen fixation. We directly quantified nodule biomass and N_2 fixation by collecting 2,334 soil cores (5.5 cm in diameter, 10 cm deep) randomly from the rooting zone beneath each individual tree using crown area to approximate root area in the rainy season of 2010 (refs 15, 39). To account for differences in rooting area with tree size, we collected 6 cores for trees with DBH of 1–5 cm, 8 cores for trees with DBH of 5–10 cm, 10 cores for trees with DBH of 10–20 cm and 12 cores for trees with DBH >20 cm, consistent with our sampling method that we developed over the past 9 years. All nodules were cut open to confirm nitrogenase activity, and total nodule biomass was expressed per soil area of each tree (g nodule m^{-2}) and per individual tree (g nodule tree^{-1}). Because natural abundance ^{15}N/^{14}N ratios cannot necessarily resolve individual tree N_2 fixation[40], we quantified nodulation from June to August 2010 and nodule mass-specific fixation using ^{15}N-tracer incubations of several species of *Inga*[15] and acetylene reduction assay (ARA) measures in 15 species of fixers across different growth conditions in Panamanian forests between 2003 and 2012 (Supplementary Table 4). We expanded these measures to the plot scale as follows: we scaled average species-specific nodule biomass (g nodule m^{-2}) to the rooting area for each individual tree and summed across all trees in each plot; for the species for which we did not directly quantify nodulation (<30% of fixer basal area), we used the average nodule biomass per area from all species. When evaluating percent nodulation across species (Fig. 2c) and the fraction of trees that were currently fixing, we considered only trees for which we had directly measured fixation.

We calculated N_2 fixation by multiplying nodule biomass by per-nodule biomass fixation rate from the two groups above, and evaluated the potential effect of species- and forest-age-specific per-nodule biomass fixation rates (discussed in Supplementary Note 3). Total ecosystem fixation rates were relatively insensitive to variations in these assumptions; we therefore report species- and forest-age-specific data on both nodulation and per-nodule biomass fixation in Table 1 and compare the alternative scenarios in Supplementary Note 3.

Model description. Our model evaluates the effect of the observed biomass accumulation trend (line in Fig. 1a; quantified from allometric measures of all individual trees across all species and all plots) on the internal forest N cycle. The model (Supplementary Box 1) tracks nitrogen mass balances (as kg N ha^{-1}) in plant pools of leaves (N_L), wood (N_W) and fine roots (N_R), as well as in plant available (N_A) and plant unavailable (N_U) forms in soil:

$$B_x = B^*_x\, t/(t^*_x + t) \quad (1)$$

$$N_L = N_R = \omega_L B_L \quad (2)$$

$$N_W = \omega_W (B_T - B_L - B_R) \quad (3)$$

$$dN_A/dt = I - \text{plant N demand} + mN_U - k(N_A) + r\,\alpha\,(\mu_L N_L(t) + \mu_R N_R(t)) \quad (4)$$

$$dN_U/dt = \beta\mu_W N_W(t) - m(N_U) + r(1-\alpha)(\mu_L N_L(t) + \mu_R N_R(t)) \quad (5)$$

Equation (1) describes the observed net plant biomass accumulation (B) over time (t) fitted as a Michaelis–Menten equation based on the tissue-specific allometric calculations outlined above (Fig. 1a and Supplementary Note 2), and where B^*_x is the equilibrium maximum biomass, t^*_x is the half-saturation constant and x refers to tissue pool (L, leaves; R, fine roots; W, wood; T, total biomass). Following standard convention[9,32,33], we assumed equal quantities of nitrogen in leaf (N_L) and fine root (N_R) biomass, and calculated both pools as leaf biomass times the fraction of nitrogen in leaf biomass (ω_L) (equation (2)). We calculated nitrogen in wood as the product of wood biomass (total plant biomass minus leaf and fine-root biomass) and the fraction of nitrogen in wood (ω_w) (equation (3)). Equation (4) gives the time-dependent change in the plant-available soil nitrogen pool, as a function of external nitrogen input (I), the plant nitrogen growth demand, the input mineralization rate (m) from plant-unavailable soil nitrogen, the loss rate (k) from the plant-available soil nitrogen pool, and the input rates of plant leaf (μ_L) and fine-root (μ_R) turnover, where r is the nitrogen fraction not retranslocated and α is the

fraction entering the soil-available pool. Plant nitrogen demand was calculated from equation (1) applied to all tissue pools and corrected for annual turnover (adjusted for retranslocation) of leaf and fine-root nitrogen. Equation (5) gives the time-dependent change in the plant-unavailable soil nitrogen pool, as a function of the input rate from wood mortality (μ_W), where β identifies the fraction of wood mortality entering the plant-unavailable pool; the loss mineralization rate (m) to the plant-available pool; and the input rates of nitrogen from leaf (μ_L) and fine-root (μ_R) turnover.

We simplified our model by externalizing all soil nitrogen with mean residence time (MRT) that exceeded the dynamics of interest in this study (taken as $>$30 years). We assumed that this soil nitrogen is large in size and originated from wood, with the flux calculated as a loss fraction ($1-\beta$) of wood mortality. We also considered a slow but constant mineralization input from this external pool, which we added to the external input term (I). We show in Supplementary Information that this simplification offered a good approximation of more complex models in which we explicitly tracked a passive soil nitrogen pool following the standard approach of CENTURY-type models.

The external nitrogen required (that is, N_2 fixation) was calculated as the difference between plant nitrogen demand and soil nitrogen supply. A positive value indicates that the internal nitrogen supply could not match the demands of the recovering forest, thus necessitating additional nitrogen inputs from symbiotic fixation. A negative balance, in turn, indicates that internal supplies were sufficient or exceeded forest demand.

Model parameterization. We parameterized the model using values reasonable for our Panamanian forests as summarized in Supplementary Table 6. We used the tissue nitrogen contents (percentage nitrogen) that we determined from our data and the literature as described above. We allowed leaves to turn over annually, which is conservative relative to our observations across the youngest forests, as faster turnover would increase external nitrogen required.

We doubled the rate of wood turnover (from 0.02 to 0.04 per year) to evaluate the effect of disturbance on fixation in mature forests (shown as dotted red line in Fig. 1c). Because we are simulating forest carbon recovery during secondary succession, we started the plant-unavailable nitrogen pool at 60% of equilibrium size; we show in Supplementary Information that increasing the pool size to 100% does not alter the dynamics of the system.

29. Hassler, S. K., Zimmerman, B., van Breugel, M., Hall, J. S. & Elsenbeer, H. Recovery of saturated hydraulic conductivity under secondary succession on former pasture in the humid tropics. *For. Ecol. Manage.* **261,** 1634–1642 (2011).
30. Neumann-Cosel, L., Zimmerman, B., Hall, J. S., van Breugel, M. & Elsenbeer, H. Soil carbon dynamics under young tropical secondary forests on former pastures—a case study from Panama. *For. Ecol. Manage.* **261,** 1625–1633 (2011).
31. Walker, L. R., Wardle, D. A., Bardgett, R. D. & Clarkson, B. D. The use of chronosequences in studies of ecological succession and soil development. *J. Ecol.* **98,** 725–736 (2010).
32. Jackson, R. B., Mooney, H. A. & Schulze, E. D. A global budget for fine root biomass, surface area, and nutrient contents. *Proc. Natl Acad. Sci. USA* **94,** 7362–7366 (1997).
33. Wolf, A., Field, C. & Berry, J. A. Allometric growth and allocation in forests: a perspective from FLUXNET. *Ecol. Appl.* **21,** 1546–1556 (2011).
34. Martin, A. R. & Thomas, S. C. A reassessment of carbon content in tropical trees. *PLoS ONE* **6,** e23533 (2011).
35. Craven, D. J. *Dynamics of Tropical Secondary Forests in Central Panama: Linking Functional Traits with Ecological Performance during Succession.* PhD thesis, Yale Univ. (2012).
36. Townsend, A. R., Cleveland, C. C., Asner, G. P. & Bustamente, M. M. C. Controls over foliar N:P ratios in tropical rain forests. *Ecology* **88,** 107–118 (2007).
37. Martius, C. Density, humidity, and nitrogen content of dominant wood species of floodplain forests (várzea) in Amazonia. *Eur. J. Wood Wood Products* **50,** 300–303 (1992).
38. Chave, J. *et al.* Towards a worldwide wood economics spectrum. *Ecol. Lett.* **12,** 351–366 (2009).
39. Muller-Landau, H. C. *et al.* Testing metabolic ecology theory for allometric scaling of tree size, growth and mortality in tropical forests. *Ecol. Lett.* **9,** 575–588 (2006).
40. Barron, A. R. *Patterns and Controls of Nitrogen Fixation in a Lowland Tropical Forest, Panama.* PhD thesis, Princeton Univ. (2007).

LETTER

OPEN

doi:10.1038/nature12511

Genome-wide signatures of convergent evolution in echolocating mammals

Joe Parker[1]*, Georgia Tsagkogeorga[1]*, James A. Cotton[1]†, Yuan Liu[2], Paolo Provero[3,4], Elia Stupka[3] & Stephen J. Rossiter[1]

Evolution is typically thought to proceed through divergence of genes, proteins and ultimately phenotypes[1–3]. However, similar traits might also evolve convergently in unrelated taxa owing to similar selection pressures[4,5]. Adaptive phenotypic convergence is widespread in nature, and recent results from several genes have suggested that this phenomenon is powerful enough to also drive recurrent evolution at the sequence level[6–9]. Where homoplasious substitutions do occur these have long been considered the result of neutral processes. However, recent studies have demonstrated that adaptive convergent sequence evolution can be detected in vertebrates using statistical methods that model parallel evolution[9,10], although the extent to which sequence convergence between genera occurs across genomes is unknown. Here we analyse genomic sequence data in mammals that have independently evolved echolocation and show that convergence is not a rare process restricted to several loci but is instead widespread, continuously distributed and commonly driven by natural selection acting on a small number of sites per locus. Systematic analyses of convergent sequence evolution in 805,053 amino acids within 2,326 orthologous coding gene sequences compared across 22 mammals (including four newly sequenced bat genomes) revealed signatures consistent with convergence in nearly 200 loci. Strong and significant support for convergence among bats and the bottlenose dolphin was seen in numerous genes linked to hearing or deafness, consistent with an involvement in echolocation. Unexpectedly, we also found convergence in many genes linked to vision: the convergent signal of many sensory genes was robustly correlated with the strength of natural selection. This first attempt to detect genome-wide convergent sequence evolution across divergent taxa reveals the phenomenon to be much more pervasive than previously recognized.

Echolocation is a complex phenotypic trait that has evolved independently in bats and whales, and which involves the production, reception and auditory processing of ultrasonic pulses for obstacle avoidance, orientation and hunting[11,12]. Recent phylogenetic studies have shown that echolocating bats are not a true group—one lineage also contains the non-echolocating Old World fruit bats (family Pteropodidae), indicating that echolocation has evolved at least twice in bats, or was lost early in the evolution of Old World fruit bats[13–15]. New evidence supports the former scenario: divergent clades of echolocating bats seem to have undergone convergent amino acid replacements in several genes implicated in hearing[8,16–18]. Furthermore, some candidate hearing genes also show parallel changes in echolocating bats and whales, again suggesting roles in high-frequency hearing[10,16,18,19]. Other genes that might function in echolocation have been identified from screens for selection in bat[20] and cetacean sequence data[21]. Here, using the evolution of echolocation as a model of phenotypic convergence, we investigated the extent to which parallel changes have occurred across the genome during the independent evolution of echolocation in bats and cetaceans.

We undertook genome sequencing of four divergent bat species, including both echolocating and non-echolocating forms. From the proposed suborder Yinpterochiroptera we sequenced the greater horseshoe bat *Rhinolophus ferrumequinum* and the greater false vampire bat *Megaderma lyra*, which exhibit 'constant frequency' (CF) and 'frequency modulated' (FM) echolocation, respectively (for details of calls, see refs 15, 22). From this suborder we also sequenced the non-echolocating straw-coloured fruit bat *Eidolon helvum*, to which we added published draft genome data from a second non-echolocating fruit bat, the large flying fox *Pteropus vampyrus*. From the second suborder, Yangochiroptera, we sequenced the CF echolocating Parnell's moustached bat *Pteronotus parnellii*, and added published data from the FM echolocating little brown bat *Myotis lucifugus*.

For each of our four focal bat species, we generated paired-end short read sequence data on a Hi-Seq 2000 platform (Illumina), assembled the raw reads *de novo* into contigs using CLC bio, and then built scaffolds in SOAPdenovo (see Methods for details). Short-read data have been deposited into the Short Read Archive under accession numbers SRR924356, SRR924359, SRR924361 and SRR924427. We conducted homology-based gene prediction and identified 20,424 genes for *R. ferrumequinum*, 20,043 for *M. lyra*, 20,455 for *E. helvum* and 20,357 for *P. parnellii*. Screening for single-copy (1-to-1) orthologous protein-coding nuclear genes conserved across eutherian mammals identified 7,612 genes present in each of our four draft genomes (see Methods for details).

To build a mammal-wide alignment of orthologous coding gene sequences (CDSs) we retrieved the CDS of each locus from 18 published mammal genomes from Ensembl (http://www.ensembl.org/, release 63), covering a broad taxonomic range and including *M. lucifugus* and *P. vampyrus* (see Methods) as well as the echolocating common bottlenose dolphin *Tursiops truncatus*. Individual gene data sets were built and aligned in frame as codons with all ambiguous sites and codons removed (see Methods). To avoid potential errors that could arise either during sequencing or data assembly, which could adversely affect phylogenetic and molecular evolution analyses, we focused on all identified genes that, after clean up, contained no missing data or gaps in any of the newly sequenced bats. In total we generated alignments for 2,326 CDSs, each spanning at least 450 base pairs, and containing a minimum of six bat species (2% of alignments had missing data from *P. vampyrus* due to its lower coverage) as well as the bottlenose dolphin and the following five other mammals: dog *Canis familiaris*, horse *Equus caballus*, cow *Bos taurus*, mouse *Mus musculus* and human *Homo sapiens*.

To detect genome-wide sequence convergence between echolocating lineages, we built an analytical pipeline based on maximum likelihood (ML) phylogenetic reconstruction[8,10,19]. In this method, we examined each amino acid along the alignment of a given CDS, and measured its fit (site-wise log-likelihood support; SSLS) to the commonly accepted species tree[23–25] (hereafter termed H_0) and to two alternative topologies in which we forced echolocating taxa into erroneous monophyletic clades representing different convergence hypotheses (see Fig. 1a; for

[1]School of Biological and Chemical Sciences, Queen Mary, University of London, London E1 4NS, UK. [2]BGI-Europe, Ole Maaløes Vej 3, DK-2200 Copenhagen N, Denmark. [3]Center for Translational Genomics and Bioinformatics, San Raffaele Scientific Institute, Via Olgettina 58, 20132 Milano, Italy. [4]Department of Molecular Biotechnology and Health Sciences, University of Turin, Via Nizza 52, I-10126 Torino, Italy. †Present address: Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.
*These authors contributed equally to this work.

details see Supplementary Fig. 1 and Methods): H_1 corresponds to all echolocating bats in a monophyletic group ('bat–bat convergence') and H_2 to all echolocating mammals together in a monophyletic group ('bat–dolphin convergence'). Using this approach we obtained the SSLS values of all amino acids under three different tree topologies. The difference in SSLS for a single site under the species tree and a given convergent tree with an identical substitution model denotes the relative support for the convergence hypothesis; for example, ΔSSLS (H_1) = SSLS (H_0) − SSLS (H_1) (where negative ΔSSLS implies support for convergence; see Supplementary Fig. 2). We quantified the extent of sequence convergence at each locus by taking the mean of its ΔSSLS values, and found 824 loci with mean support for H_1 and 392 for H_2. Using simulations we confirmed that these convergent signals were not due to neutral processes and were robust to the substitution model used (see Supplementary Methods).

We ranked the mean ΔSSLS for all 2,326 loci under both convergence hypotheses and, to assess the performance of our method, inspected the rank positions of seven hearing genes that have previously been shown to exhibit convergence and/or adaptation in echolocating mammals: prestin (*Slc26a5*), *Tmc1*, *Kcnq4* (*Kqt-4*), *Pjvk* (*Dfnb9*), otoferlin, *Pcdh15* and *Cdh23* (see Methods). Prestin was ranked 43rd (H_1) and 22nd (H_2), whereas several other loci were also ranked highly in the distribution of convergence support values (see Fig. 1b). In addition to these, we also found several other hearing genes in the top 5% supporting H_1 (*Itm2b*, *Slc4a11*) and H_2 (*Coch*, *Itm2b*, *Ercc3* and *Opa1*). Because bats and cetaceans are also known to have undergone shifts in spectral tuning and other adaptations in response to living in low light environments[26–28], we also examined the position of genes implicated in vision and found four such loci in the top 5% of genes supporting H_1 (*Lcat*, *Slc45a2*, *Rabggtb* and *Rp1*) and three supporting H_2 (*Jmjd6*, *Six* and *Rho*; see examples in Fig. 1b and Supplementary Tables 2 and 3).

We tested statistically whether the strength of sequence convergence among echolocating bats, and between echolocating bats and the bottlenose dolphin, is greater in hearing genes than in other genes (for locus selection, see Methods). For each phylogenetic hypothesis, we averaged the mean ΔSSLS values of all 21 genes in our data set that are listed as linked to either hearing and/or deafness in any taxon based on published functional annotations (see Supplementary Information). By comparing our observed values to null distributions of corresponding values obtained by randomization, we found that hearing genes had significantly more negative average values than expected by chance for bat–dolphin convergence (H_2: $z = -0.0194$, $P < 0.05$). We repeated this method for 75 genes listed as involved in vision and/or blindness, and found support, although weaker, in both cases of phenotypic convergence ($z = -0.0020$, $P \leq 0.055$ and $z = -0.0097$, $P \leq 0.09$). Loci previously reported to have association with echolocation had strong support by randomization for both hypotheses ($P \leq 0.01$ in both cases).

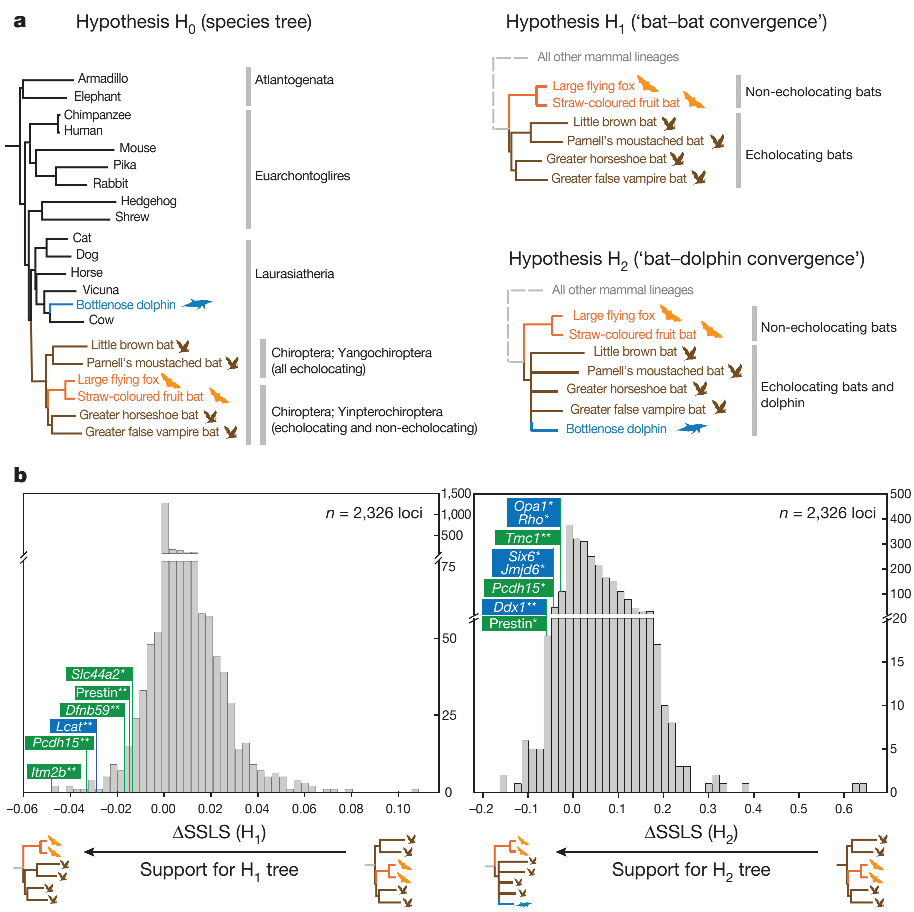


Figure 1 | Convergence hypotheses and genomic distribution of support. **a**, For each locus, the goodness-of-fit of three separate phylogenetic hypotheses was considered: (left) H_0, the accepted species phylogeny based on recent findings (for example, refs 14, 23–25); (top-right panel) H_1, or 'bat–bat convergence', in which echolocating bat lineages (shown in brown) are forced to form a monophyletic group to the exclusion of non-echolocating Old World fruit bats (shown in orange); and (bottom-right panel) H_2, or 'bat–dolphin convergence', in which the echolocating bat lineages and the dolphin (blue) form a monophyletic group to the exclusion of all non-echolocating mammals. See Methods for details of model fitting and topologies. **b**, The distribution of convergence signal across 2,326 loci in 14–22 representative mammalian taxa, as measured by locus-wise mean site-specific likelihood support for the species topology (H_0) over (left) the 'bat–bat' hypothesis uniting echolocating bats (that is, ΔSSLS (H_1)) and (right) bat–dolphin hypothesis (that is, ΔSSLS (H_2)). Representative hearing and vision loci are shown in green and blue, respectively; for each locus significance levels based on simulation denote whether it had significant counts of convergent sites after correcting for expected counts in random (control) phylogenies (*), and additionally whether strength of positive selection (dN/dS) and convergence (ΔSSLS) at sites under selection in echolocators were correlated (**); see Supplementary Table 4 and Methods.

We inspected direct protein–protein interactions among the 117 genes (top 5%) for H_2 in networks of published interactions and found that 17 of these loci formed a single large network of direct interactions (for details see Supplementary Information and Supplementary Fig. 9) centred around *Tp63* (*p63*) and *Cdk1* (*p34*); *p63* has been shown recently to be involved in cochlea development[29] and *Cdk1* has been shown to be important for development and regeneration of hair cells in the inner ear[30]. We also scanned databases of tissue-specific RNA expression in humans, finding some genes with elevated hypothalamus expression in human orthologues (Supplementary Information and Supplementary Table 13).

Most of the loci supporting the monophyly of echolocating bats, or the clade of echolocating bats plus dolphin, have no known roles in the sensory perception of sound or light. Yet given that many of these loci encode proteins with poorly characterized functions, a role in hearing or vision cannot be ruled out and in this respect it is noteworthy that highly ranked genes for H_1 (top 5%) included five solute carrier proteins related to prestin, a motor protein that drives the cochlear amplifier[8,19] (see Supplementary Table 2). Other loci that seem to support convergence among echolocating taxa can be more confidently ruled out as having roles in sensory perception; however, some of these loci may instead be associated with phenotypic traits that are correlated with aspects of echolocation.

Sequence convergence in loci related to shared phenotypic traits provides compelling but indirect evidence that genome-wide examples of convergence reported here are likely to be due to adaptive selection rather than neutral evolution. Analysis of the distribution of support for convergence within each locus indicated that mean support is driven by a few sites per locus (see Supplementary Information). For echolocating taxa under each convergent scenario, we directly estimated the strength of selection at each site as the ratio (omega, ω) of the rate of non-synonymous substitutions (dN) to the rate of synonymous substitutions (dS; $\omega > 1$ being indicative of molecular adaptation; see Methods). To test whether site-wise support for a given hypothesis is driven by selection, we correlated absolute site-wise ΔSSLS and site-wise ω and found a strong relationship after correcting for locus identity (H_1, $P = 0.0336$; H_2, $P < 0.001$).

Previous work on prestin suggested that sequence convergence among echolocating taxa was a consequence of positive selection[19]. To determine the extent to which adaptive convergence occurs across the genome, for each locus we fitted the linear relationship between ΔSSLS and corresponding site-wise ω in echolocating lineages for those sites that show different selection pressures between echolocators and other taxa. We proposed that loci with sites undergoing diversifying selection for convergence (adaptive convergence) would exhibit a negative correlation, whereas those loci with sites undergoing divergence driven by diversifying selection, but under the accepted species topology (adaptive divergence), would exhibit a positive correlation. On the basis of whether the 95% confidence interval of the slope was below zero, for H_1 we classified 92 loci as putatively adaptive convergent; conversely, in 111 loci, the 95% confidence interval of the slope was greater than zero, indicating adaptive divergence. Using the same approach for H_2, we classified 59 loci as adaptive convergent and 212 loci as adaptive divergent (Fig. 2). The larger number of adaptive divergent loci seen under H_2 reflects the stronger average relative support for H_0 across the data set in H_0H_2 comparisons, as opposed to H_0H_1 comparisons; this is to be expected because H_2 is a more radical rearrangement of the species phylogeny.

To explore more fully the relationship between signatures of convergence and adaptive evolution, we predicted the level of convergent evolution expected (ΔSSLS) after prolonged diversifying selection ($\omega = 2$) using the fitted regression models for each locus. Genes associated with hearing and vision (including those encoding the solute carriers

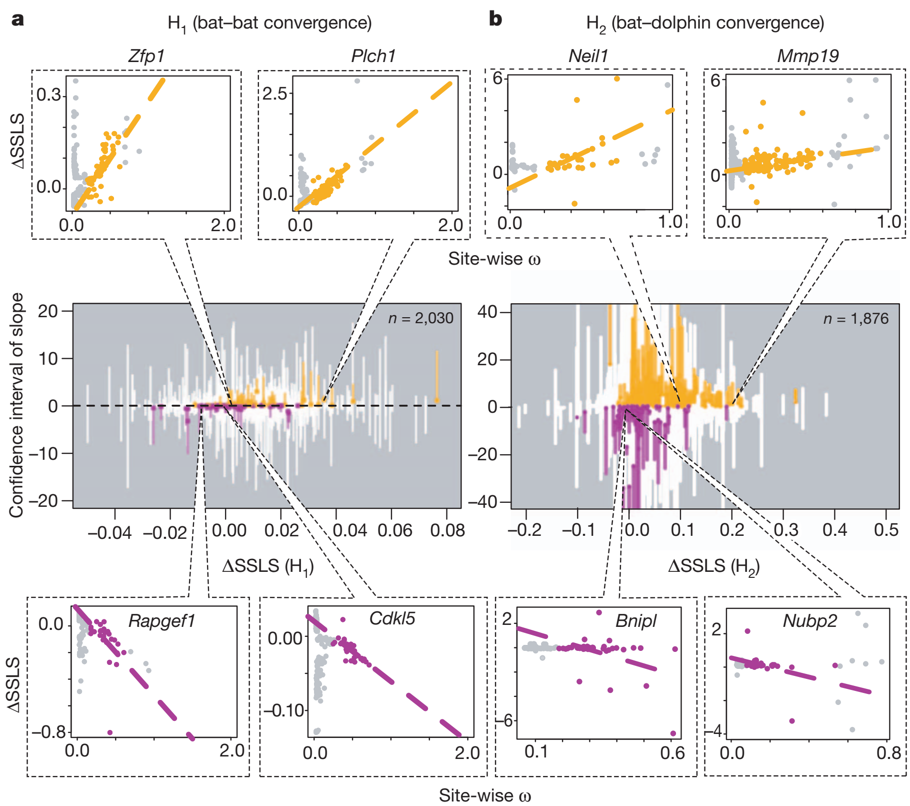


Figure 2 | Relationship between strength of convergence signal and adaptive selection. **a**, **b**, For hypotheses H_1(**a**) and H_2(**b**) ($n = 2{,}030$ and $1{,}876$ loci, respectively), the 95% confidence intervals of the coefficient (slope) for locus-wise regressions between site-wise support for convergence and site-wise ω for sites under diversifying selection are plotted. In each plot, loci showing a negative relationship, as characterized by a slope significantly below zero, are consistent with an evolutionary trajectory of adaptive convergence (purple line, with filled circle indicating upper 95% limit) and loci showing a positive relationship, with a slope of greater than zero, are consistent with an evolutionary trajectory of adaptive divergence (orange line, with filled circle indicating lower 95% limit). Insets show two examples of adaptive convergence and divergence under each hypothesis. Full details of ω estimation and regression fitting are given in the Methods.

Slc44a2 and Slc4a11, associated with deafness, and the integrin Itgal, associated with blindness) showed significantly stronger convergence in these models than the background loci (see Supplementary Information). The genes identified in this analysis show a pattern of substitutions under differing selection regimes in echolocating and non-echolocating lineages that tend to have high ω and be convergent in echolocating lineages; we therefore suggest that sustained adaptive selection in these loci is likely to reinforce the signal for convergence as parallel changes accrue, in contrast to neutrally driven or adaptive divergence, which would tend to reinforce support for the species phylogeny over time.

Our genome-wide analysis shows that natural selection has acted on three echolocating lineages (cetaceans and two separate bat lineages) to produce a complex pattern of changes in protein sequence, including both divergent and, more surprisingly, extensive convergent changes. Many of these changes are in genes that may be associated with the shift in primary sensory modality (between vision and echolocation), either directly or through the associated complex changes in ecology and natural history. Furthermore, this work identifies clear targets for future experimental work, for instance tissue-specific RNA expression analyses. This study represents the first systematic attempt to provide a framework for the genomic analysis of sequence convergence associated with independently shared phenotypes. Our findings strongly suggest that, despite many recent papers reporting sequence convergence in particular candidate genes, the importance of this mode of molecular evolutionary change is relatively underappreciated, and is under-exploited in seeking to understand the genetic basis of complex traits such as echolocation.

METHODS SUMMARY

Taxonomic coverage, sequencing and assembly. Genome-wide short-read sequences (Hi-Seq, 500-bp inserts) were generated and assembled for one non-echolocating (*Eidolon helvum*) and three echolocating bats (*Rhinolophus ferrumequinum*, *Megaderma lyra* and *Pteronotus parnellii*). The former three belong to the suborder Yinpterochiroptera and the latter to the Yangochiroptera. A total of 7,612 coding sequences showed 1-to-1 orthology with *Homo* sequences and were present in all four taxa. Of these, 2,326 loci were aligned in MAFFT with published data from 17 additional taxa, including the non-echolocating bat *Pteropus vampyrus*, echolocating bat *Myotis lucifugus* and echolocating dolphin *Tursiops truncatus*.

Convergence pipeline. We built a pipeline to estimate the strength of natural selection (dN/dS ratio, ω) and support for convergence (ΔSSLS) for each amino acid by maximum likelihood. ω values were estimated using published software; ΔSSLS was calculated as the difference in fitted site-wise log-likelihoods under two competing hypotheses: a 'null' phylogenetic tree, reflecting the consensus species tree, and one of two 'alternative' phylogenies, in which either all echolocating bats, or all echolocating taxa (echolcating bats plus dolphin), were artificially forced into a monophyletic clade. Ancestral amino acid states were inferred using parsimony.

Null distributions. We compared observed ΔSSLS values to ΔSSLS estimated from (1) neutrally evolving amino acids simulated by Markov chain Monte Carlo mixture models ($n \geq 1,000$); and (2) random 'alternative' topologies ($n = 100$).

Locus-wise adaptation/convergence regressions. For each locus the relationship between selection and convergence was modelled using least-squares regression. 95% confidence intervals were calculated for the line equation parameters. The fitted line equations were extrapolated to model the convergence signal under strong adaptive selection ($\omega = 2$).

Hearing and vision genes. We used the gene ontology database DAVID (http://david.abcc.ncifcrf.gov) to identify which of our loci were putatively associated with hearing or vision. Simulation by randomization ($n = 1,000$) was used to compare the observed mean convergence values among these sensory loci to null distributions.

Full Methods and any associated references are available in the online version of the paper.

Received 21 January; accepted 30 July 2013.

Published online 4 September 2013.

1. Soskine, M. & Tawfik, D. S. Mutational effects and the evolution of new protein functions. *Nature Rev. Genet.* **11,** 572–582 (2010).
2. Clark, A. G. *et al.* Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450,** 203–218 (2007).
3. Hughes, J. F. *et al.* Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* **463,** 536–539 (2010).
4. Hoy, R. R. Evolution. Convergent evolution of hearing. *Science* **338,** 894–895 (2012).
5. Grant, P. R., Grant, B. R., Markert, J. A., Keller, L. F. & Petren, K. Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution* **58,** 1588–1599 (2004).
6. Zhang, J. Z. & Kumar, S. Detection of convergent and parallel evolution at the amino acid sequence level. *Mol. Biol. Evol.* **14,** 527–536 (1997).
7. Kriener, K., O'hUigin, C., Tichy, H. & Klein, J. Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. *Immunogenetics* **51,** 169–178 (2000).
8. Li, G. *et al.* The hearing gene Prestin reunites echolocating bats. *Proc. Natl Acad. Sci. USA* **105,** 13959–13964 (2008).
9. Castoe, T. A. *et al.* Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc. Natl Acad. Sci. USA* **106,** 8986–8991 (2009).
10. Liu, Y., Rossiter, S. J., Han, X., Cotton, J. A. & Zhang, S. Cetaceans on a molecular fast track to ultrasonic hearing. *Curr. Biol.* **20,** 1834–1839 (2010).
11. Vater, M. & Kössl, M. in *Echolocation in Bats and Dolphins* (eds Thomas, J. T., Moss, C. F. & Vater, M.) 89–98 (Univ. Chicago Press, 2004).
12. Au, W. W. L. & Simmons, J. A. Echolocation in dolphins and bats. *Phys.Today* **60,** 40–45 (2007).
13. Teeling, E. C. *et al.* Microbat paraphyly and the convergent evolution of a key innovation in Old World rhinolophoid microbats. *Proc. Natl Acad. Sci. USA* **99,** 1431–1436 (2002).
14. Teeling, E. C. *et al.* Molecular evidence regarding the origin of echolocation and flight in bats. *Nature* **403,** 188–192 (2000).
15. Jones, G. & Holderied, M. W. Bat echolocation calls: adaptation and convergent evolution. *Proc. R. Soc. B* **274,** 905–912 (2007).
16. Davies, K. T. J., Cotton, J. A., Kirwan, J. D., Teeling, E. C. & Rossiter, S. J. Parallel signatures of sequence evolution among hearing genes in echolocating mammals: an emerging model of genetic convergence. *Heredity* **108,** 480–489 (2012).
17. Liu, Y. *et al.* The voltage-gated potassium channel subfamily KQT member 4 (KCNQ4) displays parallel evolution in echolocating bats. *Mol. Biol. Evol.* **29,** 1441–1450 (2012).
18. Shen, Y.-Y., Liang, L., Li, G.-S., Murphy, R. W. & Zhang, Y.-P. Parallel evolution of auditory genes for echolocation in bats and toothed whales. *PLoS Genet.* **8,** e1002788 (2012).
19. Liu, Y. *et al.* Convergent sequence evolution between echolocating bats and dolphins. *Curr.Biol.* **20,** R53–R54 (2010).
20. Zhang, G. *et al.* Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* **339,** 456–460 (2013).
21. Sun, Y.-B. *et al.* Genome-wide scans for candidate genes involved in the aquatic adaptation of dolphins. *Genome Biol. Evol.* **5,** 130–139 (2013).
22. Jones, G. & Teeling, E. C. The evolution of echolocation in bats. *Trends Ecol. Evol.* **21,** 149–156 (2006).
23. Murphy, W. J., Pringle, T. H., Crider, T. A., Springer, M. S. & Miller, W. Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Res.* **17,** 413–421 (2007).
24. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478,** 476–482 (2011).
25. Zhou, X. *et al.* Phylogenomic analysis resolves the interordinal relationships and rapid diversification of the laurasiatherian mammals. *Syst. Biol.* **61,** 150–164 (2012).
26. Zhao, H. B. *et al.* The evolution of color vision in nocturnal mammals. *Proc. Natl Acad. Sci. USA* **106,** 8980–8985 (2009).
27. Zhao, H. B. *et al.* Rhodopsin molecular evolution in mammals inhabiting low light environments. *PLoS ONE* **4,** e8326 (2009).
28. Fasick, J. I. & Robinson, P. R. Spectral-tuning mechanisms of marine mammal rhodopsins and correlations with foraging depth. *Vis. Neurosci.* **17,** 781–788 (2000).
29. Terrinoni, A. *et al.* Role of p63 and the Notch pathway in cochlea development and sensorineural deafness. *Proc. Natl. Acad. Sci. USA* **110,** 7300–7305 (2013).
30. Ryan, A. F. The cell cycle and the development and regeneration of hair cells. *Curr. Top. Dev. Biol.* **57,** 449–466 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to K. Baker, L. Davalos, D. Hayman, E. Koilmani, Y. Liu, A. Peel, R. Ransome and A. Rodriguez for providing material for sequencing. We thank C. Walker (Queen Mary GridPP High Throughput Cluster) and A. Terry and C. Mein (Barts and the London Genome Centre) for providing access to computing facilities, and for assistance with running analyses. We are also grateful to S. Dodsworth, R. Buggs, K. Davies, J. Kirkpatrick, R. Nichols, Y. Wurm and S. Young for comments on the manuscript. This work was funded by Biotechnology and Biological Sciences Research grant BB/H017178/1 awarded to S.J.R., E.S. and J.A.C.

Author Contributions S.J.R. conceived the study and secured funding together with J.A.C. and E.S. J.P. conducted all phylogenetic, convergence and selection analyses with input from S.J.R., G.T. and J.A.C. Processing and analyses of sequence data was undertaken by G.T., with input from E.S., who also conducted gene ontology analyses with P.P. Raw sequence data was generated under direction of Y.L. and S.J.R. The paper was written and figures prepared by J.P. and S.J.R. with input from G.T., J.A.C. and E.S.

Author Information Short-read data have been deposited into the Short Read Archive under accession numbers SRR924356, SRR924359, SRR924361 and SRR924427. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.P. (j.d.parker@qmul.ac.uk) or S.J.R. (s.j.rossiter@qmul.ac.uk).

LETTER

doi:10.1038/nature12495

Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy

Alejandra Sanjuan-Pla[1,2], Iain C. Macaulay[3,4], Christina T. Jensen[3,4], Petter S. Woll[3,4], Tiago C. Luis[3,4], Adam Mead[3,4], Susan Moore[1,2], Cintia Carella[2], Sahoko Matsuoka[3,4], Tiphaine Bouriez Jones[3,4], Onima Chowdhury[3,4], Laura Stenson[3,4], Michael Lutteropp[3,4], Joanna C. A. Green[3,4], Raffaella Facchini[3,4], Hanane Boukarabila[3,4], Amit Grover[3], Adriana Gambardella[3], Supat Thongjuea[3], Joana Carrelha[3,4], Paul Tarrant[3,4], Deborah Atkinson[3,4], Sally-Ann Clark[3,4], Claus Nerlov[1,2,3]* & Sten Eirik W. Jacobsen[3,4]*

The blood system is maintained by a small pool of haematopoietic stem cells (HSCs), which are required and sufficient for replenishing all human blood cell lineages at millions of cells per second throughout life. Megakaryocytes in the bone marrow are responsible for the continuous production of platelets in the blood, crucial for preventing bleeding—a common and life-threatening side effect of many cancer therapies—and major efforts are focused at identifying the most suitable cellular and molecular targets to enhance platelet production after bone marrow transplantation or chemotherapy[1]. Although it has become clear that distinct HSC subsets exist that are stably biased towards the generation of lymphoid or myeloid blood cells[2–4], we are yet to learn whether other types of lineage-biased HSC exist or understand their inter-relationships and how differently lineage-biased HSCs are generated and maintained. The functional relevance of notable phenotypic and molecular similarities between megakaryocytes and bone marrow cells with an HSC cell-surface phenotype[5–8] remains unclear. Here we identify and prospectively isolate a molecularly and functionally distinct mouse HSC subset primed for platelet-specific gene expression, with enhanced propensity for short- and long-term reconstitution of platelets. Maintenance of platelet-biased HSCs crucially depends on thrombopoietin, the primary extrinsic regulator of platelet development[9]. Platelet-primed HSCs also frequently have a long-term myeloid lineage bias, can self-renew and give rise to lymphoid-biased HSCs. These findings show that HSC subtypes can be organized into a cellular hierarchy, with platelet-primed HSCs at the apex. They also demonstrate that molecular and functional priming for platelet development initiates already in a distinct HSC population. The identification of a platelet-primed HSC population should enable the rational design of therapies enhancing platelet output.

One of the megakaryocyte genes most highly expressed in mouse HSC-enriched bone marrow cells is that encoding von Willebrand factor (vWF)[7,10], a protein involved in platelet aggregation[11] and otherwise exclusively expressed on platelets and endothelium. However, single-cell analysis showed that only a fraction of phenotypically defined HSCs expressed *Vwf*[7], suggesting that platelet/megakaryocyte-related genes may be expressed in only a distinct HSC subpopulation or in contaminating megakaryocyte progenitors. To determine whether a megakaryocyte-primed HSC subset exists, we generated a bacterial artificial chromosome (BAC) transgenic *Vwf-eGFP* reporter that expresses enhanced green fluorescent protein (eGFP; Extended Data Fig. 1a), and verified that its expression in mature cells and progenitors was restricted to endothelial cells, platelets and megakaryocyte-lineage progenitors (Extended Data Figs 1 and 2a–d). Around 60% of HSCs, stringently defined as Lin$^-$ Sca-1$^+$ c-Kit$^+$ (LSK) CD150$^+$ CD48$^-$ CD34$^-$ (refs 12, 13), were distinctly vWF–eGFP$^+$ (Fig. 1a). By contrast, lymphoid-primed multipotent progenitors (Extended Data Fig. 2e), the most primitive multipotent progenitor lacking megakaryocyte but sustaining myeloid lineage potential[6,7], lacked *Vwf-eGFP* expression. *Vwf-eGFP* expression reflected endogenous *Vwf* expression in HSCs, not affected by the transgene (Extended Data Fig. 3a, b). *Vwf-eGFP* expression also sub-fractionated HSCs defined by lack of CD49b expression[14] or CD201expression[15] (Extended Data Fig. 3c), and different levels of CD150 (ref. 16; Fig. 1b), CD41 (ref. 17) or CD86 expression[18] (Extended Data Fig. 3d). Both vWF–eGFP$^+$ and vWF–eGFP$^-$ HSCs showed high Hoechst efflux (Extended Data Fig. 3e) and were highly quiescent (Extended Data Fig. 3f), properties characteristic of true HSCs[19]. Single vWF–eGFP$^+$ LSK CD150$^+$ CD48$^-$ cells formed large megakaryocyte-containing mixed-lineage colonies (Extended Data Fig. 3g). *Vwf-eGFP* expression therefore divides multipotent HSCs, homogeneous by thus far characterized phenotypes, into two prospectively isolatable subsets (henceforth designated vWF$^+$ and vWF$^-$ HSCs).

To establish whether differential expression of *vWF-eGFP* discriminates between functionally distinct HSC subsets, we performed competitive transplantation of purified vWF$^+$ and vWF$^-$ HSCs (Extended Data Figs 4 and 5). Although not statistically significant, a tendency towards higher contribution of myeloid and lower B- and T-cell lineages of the immune system was seen when transplanting 50 vWF$^+$ compared to 50 vWF$^-$ HSCs (Fig. 1c–e). More notably, vWF$^+$ HSCs consistently, in the short and long term, generated higher platelet output than vWF$^-$ HSCs (Fig. 1f). This platelet bias was maintained for the 32-week monitoring and sustained after serial transplantation (Extended Data Fig. 4c), as was the tendency towards higher myeloid and lower lymphoid reconstitution (Extended Data Fig. 4c), suggesting that platelet bias is stably and intrinsically associated with long-term vWF$^+$ HSC activity.

To investigate the lineage bias of vWF$^+$ HSCs at the clonal level, we next transplanted more limited numbers (10 cells) of vWF$^-$ and vWF$^+$ HSCs, resulting in only 25–30% of mice being long-term reconstituted (Table 1), and compatible with around 80% of recipients being repopulated by single HSCs (Poisson distribution). Notably, all mice long-term reconstituted with limiting numbers of vWF$^-$ HSCs showed clear lymphoid-biased reconstitution in peripheral blood, whereas vWF$^+$ HSCs typically displayed platelet-biased or platelet/myeloid-biased long-term reconstitution, although in a few instances lymphoid bias was observed (Table 1, Fig. 1g–j and Extended Data Fig. 6a–c). To obtain more mice reconstituted with single HSCs, we transplanted mice with five as well as single vWF$^+$ HSCs. All four mice reconstituted after transplantation of five vWF$^+$ cells showed long-term platelet-biased or

[1]Institute for Stem Cell Research and MRC Centre for Regenerative Medicine, University of Edinburgh, Edinburgh EH9 16UU, UK. [2]EMBL Mouse Biology Program, Monterotondo 00015, Italy. [3]MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK. [4]Haematopoietic Stem Cell Biology Laboratory, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK.
*These authors contributed equally to this work.

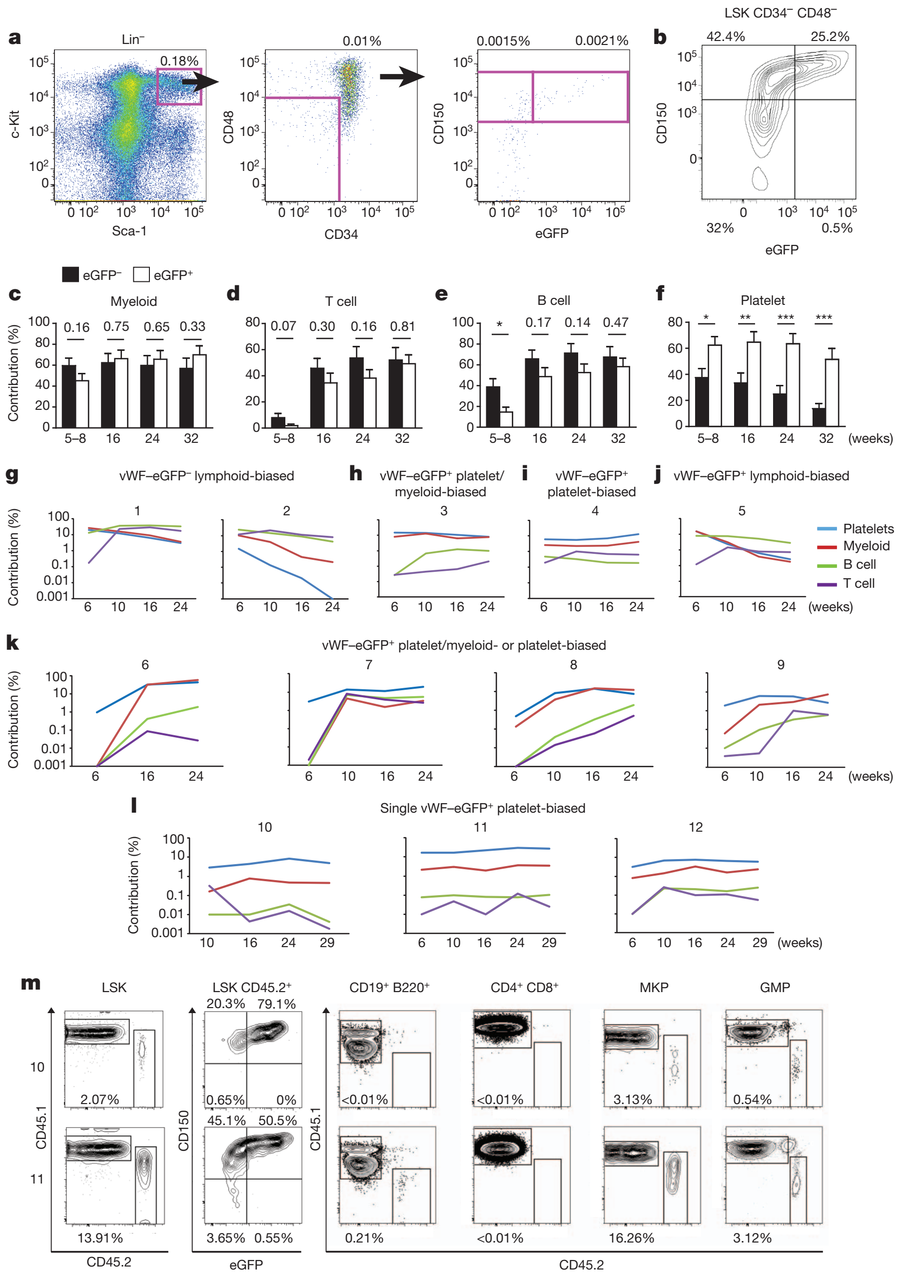


Figure 1 | vWF–eGFP-expressing platelet-biased haematopoietic stem cells. **a**, vWF–eGFP expression in adult LSK CD150$^+$ CD48$^-$ CD34$^-$ HSCs ($n = 8$). **b**, Co-expression in HSCs of vWF–eGFP with CD150 ($n = 3$). **c**–**f**, Mean (± s.e.m.) contribution of 50 vWF$^+$ and vWF$^-$ LSK CD34$^-$ CD150$^+$ CD48$^-$ HSCs to peripheral blood myeloid cells (**c**), T cells (**d**), B cells (**e**) and platelets (**f**) (3 experiments, $n = 15$). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ (Student's t-test). **g**–**l**, Representative peripheral blood reconstitution kinetics in mice transplanted with 10 vWF$^-$ LSK CD34$^-$ CD150$^+$ CD48$^-$ HSCs (**g**), 10 vWF$^+$ HSCs (**h**–**j**), 5 vWF$^+$ HSCs (**k**) and 1 single vWF$^+$ HSC (**l**). **m**, Contribution to eGFP$^+$ and eGFP$^-$ HSC compartment and different progenitors 31 weeks after transplantation in mice from **l**. GMP, granulocyte/macrophage progenitors; MKP, megakaryocyte progenitors.

Table 1 | Lineage reconstitution patterns in transplanted mice

	Platelet/myeloid-biased				Lymphoid-biased	Reconstituted/ total mice
	P	P/M	M	Other		
GFP$^-$ recipients (n)	0	0	0	0	10	10/42
Platelet (%)					0.77 (0.63)	
Myeloid (%)					1.22 (0.91)	
B cell (%)					12.89 (4.98)	
T cell (%)					10.94 (3.86)	
GFP$^+$ recipients (n)	5	4	1	2	3	15/48
Platelet (%)	4.39 (1.18)	13.99 (3.60)	4.50	2.77 (1.50)	16.35 (15.96)	
Myeloid (%)	0.65 (0.45)	16.19 (5.98)	7.56	1.10 (0.54)	17.50 (17.05)	
B cell (%)	0.04 (004)	4.81 (2.78)	0.92	6.83 (0.68)	34.10 (26.28)	
T cell (%)	0.18 (0.13)	1.04 (0.70)	0.14	3.44 (2.34)	22.79 (11.87)	

Number (n) of mice with different patterns/bias of lineage reconstitution, and total frequency of reconstituted mice, 16 weeks after transplantation of 10 vWF–eGFP$^+$ or vWF–eGFP$^-$ HSCs. Mean values are shown, with s.e.m. in parentheses. Mice were defined as reconstituted with transplanted HSCs if >1% contribution towards at least one of the platelet, myeloid or B-cell lineages at two consecutive time points (10 and 16 weeks). The reconstitution patterns were (1) platelet-biased: HSC contribution to platelets >50% higher than myeloid and B cells at 10 and 16 weeks; (2) platelet/myeloid-biased: contribution to platelets and myeloid cells <50% higher than each other, but >50% higher than B cells at 10 and 16 weeks; (3) myeloid-biased: contribution to myeloid >50% higher than platelets and B cells at 10 and 16 weeks; (4) other: lineage bias different at 10 and 16 weeks or balanced (<50% difference); and (5) lymphoid-biased: contribution to B cells >50% higher than platelets and myeloid cells at 10 and 16 weeks. Also shown is mean (s.e.m.) test-cell contribution to each lineage, for each category of lineage bias. M, myeloid-biased; P, platelet-biased; P/M, platelet/myeloid-biased.

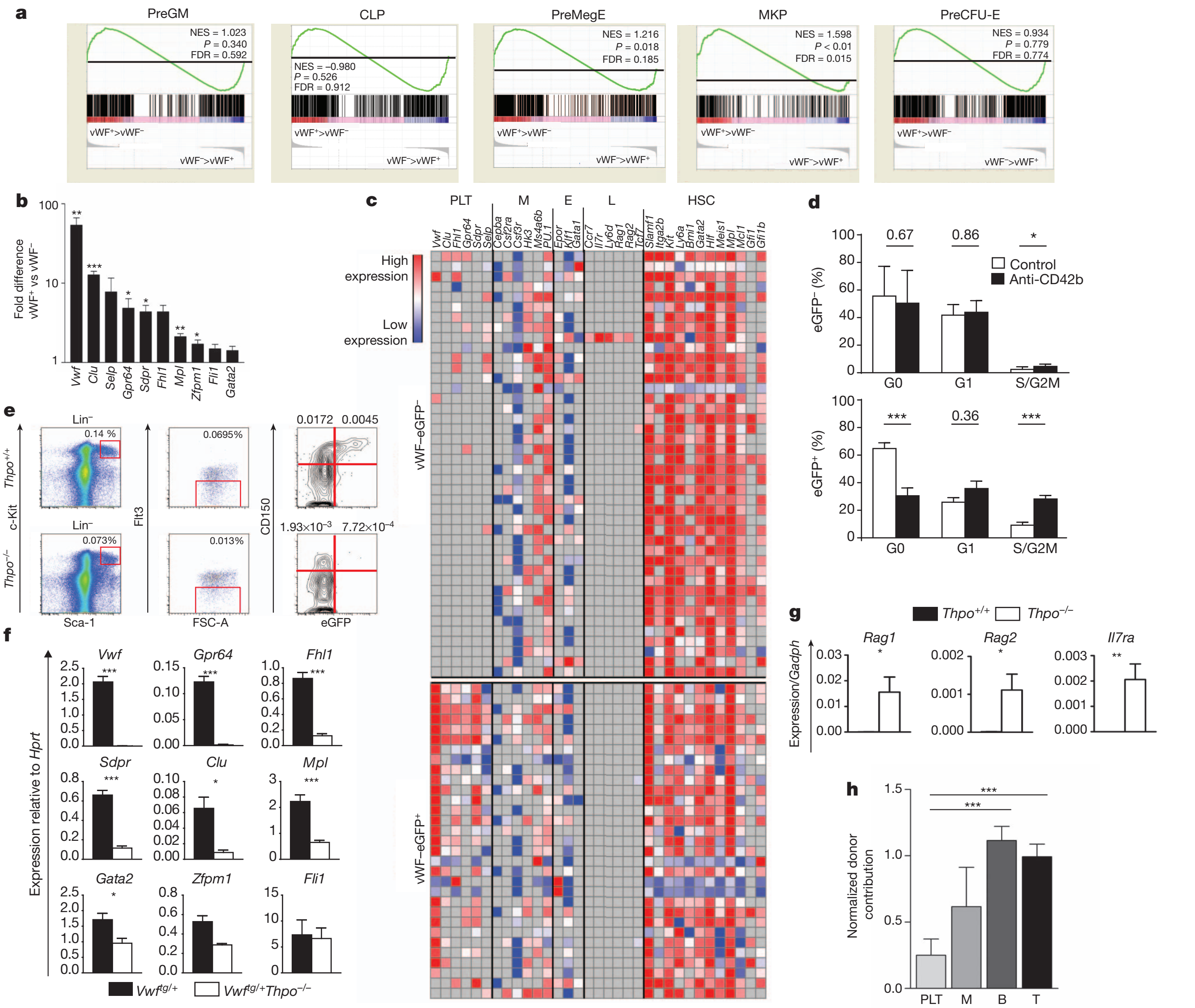


Figure 2 | THPO regulates platelet-primed vWF–eGFP⁺ HSCs. a, GSEA of indicated signatures comparing vWF$^+$ and vWF$^-$ HSCs. CLP, common lymphoid progenitors; FDR, false discovery rate; NES, normalized enrichment score; preCFU-E, pre-erythrocyte colony-forming units; preGM, pre-granulocyte/macrophage progenitors; preMegE, pre-megakaryocyte/erythrocyte progenitors. **b**, Fold difference of megakaryocyte gene expression between vWF$^+$ and vWF$^-$ HSCs ($n = 5$ biological replicates). **c**, Representative heatmap (one of three experimental plates) for expression of lineage-affiliated genes in single vWF–eGFP$^-$ ($n = 119$) and vWF–eGFP$^+$ ($n = 111$) HSCs. *PU.1* is also known as *Sfpi1*. E, erythroid; L, lymphoid; M, myeloid; PLT, platelet. **d**, Mean cell cycle phase distribution of vWF$^-$ (top) and vWF$^+$ (bottom) HSCs 24 h after anti-CD42b-mediated platelet depletion ($n = 8$). **e**, Mean vWF-eGFP expression in LSK Flt3$^-$ CD150$^+$ cells from 15-week-old *Thpo*$^{+/+}$;*Vwf-eGFP*$^{tg/+}$ and *Thpo*$^{-/-}$;*Vwf-eGFP*$^{tg/+}$ mice ($n = 3$). FSC-A, forward scatter area. **f**, Mean (± s.e.m.) megakaryocyte gene expression in *Thpo*$^{+/+}$ and *Thpo*$^{-/-}$ LSK CD34$^-$ Flt3$^-$ bone marrow cells of six replicate determinations. **g**, Lymphoid gene expression in vWF$^-$ LSK CD150$^+$ Flt3$^-$ cells from 15-week-old *Thpo*$^{+/+}$;*Vwf-eGFP*$^{tg/+}$ and *Thpo*$^{-/-}$;*Vwf-eGFP*$^{tg/+}$ mice ($n = 3$). **h**, Peripheral blood reconstitution analysis 16 weeks after transplantation of *Thpo*$^{-/-}$ *Vwf-eGFP*$^{tg/+}$ unfractionated bone marrow cells (CD45.1). For each individual mouse, donor contribution to each lineage was normalized to corresponding CD45.1$^+$ chimaerism ($n = 7$). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ (Student's *t*-test)

platelet/myeloid-biased reconstitution (Fig. 1k and Extended Data Fig. 6d), and all mice (3 out of 17 transplanted) reconstituted with a single vWF$^+$ HSC showed prominent and stable platelet-biased reconstitution, with barely detectable B- and in particular T-cell contribution at any time (Fig. 1l and Extended Data Fig. 6e). Notably, virtually every mouse that reconstituted with limiting numbers (1, 5 or 10) of vWF$^+$ HSCs, whether long-term reconstituted with a platelet or platelet/myeloid bias, showed a very distinct and more exclusive platelet bias (rather than platelet/myeloid bias) at the earliest reconstitution analysis (6 weeks), not seen with vWF$^-$ HSCs (Fig. 1g–l). The distinct long-term lineage-biased reconstitution patterns of vWF$^-$ and vWF$^+$ HSCs in peripheral blood were confirmed at the bone marrow progenitor cell level (Fig. 1m and Extended Data Fig. 7).

Peripheral blood lineage and bone marrow progenitor analyses of mice reconstituted with single vWF$^+$ HSCs establish the existence of strongly and stably platelet-biased HSCs. Furthermore, they show that vWF$^+$ platelet-biased HSCs typically also contribute considerably to other myeloid lineages, whereas their lymphoid, in particular T-cell, contribution is usually very marginal (Table 1, Fig. 1l and Extended Data Fig. 6e), implicating the existence of platelet-primed HSCs with an almost exclusively platelet/myeloid-restricted reconstitution potential. Stable platelet-biased output was associated with the maintenance of a high proportion of LSK CD150$^+$ vWF$^+$ HSCs, whereas lymphoid-biased output generated from either vWF$^+$ or vWF$^-$ HSCs was associated with engraftment of a vWF$^-$ HSC population (Fig. 1m and Extended Data Fig. 7).

Affymetrix gene profiling of vWF^{-} and vWF^{+} HSCs (Supplementary Table 6) showed higher expression of megakaryocyte-lineage genes (*Selp*, *Sdpr*, *Gpr64* and *Clu*) in vWF^{+} HSCs. Accordingly, gene-set enrichment analysis (GSEA) using gene sets from lineage-restricted progenitors[20] revealed no bias in myeloid or lymphoid programming between vWF^{+} and vWF^{-} HSCs, whereas genes associated with bipotent megakaryocyte/erythroid progenitors were enriched in vWF^{+} HSCs (Fig. 2a). Separation of erythroid and megakaryocyte programs revealed a specific and highly significant enrichment of megakaryocyte-associated genes in vWF^{+} HSCs (Fig. 2a), results confirmed by quantitative PCR (qPCR; Fig. 2b). Multiplex single-cell qPCR (Fig. 2c) demonstrated that vWF^{+} (and not vWF^{-}) LSK CD34^{-} CD150^{+} CD48^{-} HSCs are primed at the single-cell level for multiple megakaryocyte-affiliated genes, while co-expressing myeloid and erythroid, but not lymphoid, genes. vWF^{+} HSCs are therefore a distinct, molecularly and functionally megakaryocyte-primed and platelet-biased HSC population.

Thrombopoietin (THPO), crucial for megakaryocyte and platelet production[9], is also important for maintaining HSCs[21,22]. *Mpl*, which encodes the THPO receptor, was expressed at higher levels in vWF^{+} than in vWF^{-} HSCs (Fig. 2b). After acute platelet depletion, resulting in increased THPO levels (Extended Data Fig. 8a–c), otherwise quiescent vWF^{+} HSCs were rapidly and selectively recruited into active cell cycle (Fig. 2d and Extended Data Fig. 8d). In adult *Vwf-eGFP* mice lacking *Thpo* (*Thpo*$^{-/-}$; *Vwf-eGFP*$^{tg/+}$ mice), vWF–eGFP^{+} HSCs were markedly and selectively reduced (Fig. 2e). Loss of vWF–eGFP^{+} HSCs was preceded by selective depletion of megakaryocyte-lineage gene expression in HSCs (Fig. 2f), whereas myeloid, erythroid and some lymphoid gene expression was maintained (Extended Data Fig. 9a). However, several lymphoid genes were upregulated in vWF–eGFP^{-} HSCs from THPO-deficient mice (Fig. 2g), and THPO-deficient bone marrow cells gave lymphoid-biased reconstitution when transplanted (Fig. 2h and Extended Data Fig. 9b). After re-exposure to THPO in wild-type recipients, THPO-deficient bone marrow cells reconstituted vWF^{-} as well as vWF^{+} HSCs (Extended Data Fig. 9c), further supporting the idea that THPO is crucially required for maintenance of platelet-primed HSCs.

In addition to vWF^{-} HSCs giving lymphoid-biased reconstitution, we observed that the LSK CD150^{+} CD48^{-} vWF^{+} HSC compartment was long-term reconstituted in recipients of vWF^{+}, but not vWF^{-}, HSCs, whereas both HSC populations regenerated LSK CD150^{+} CD48^{-} vWF^{-} HSCs (Figs 1m and 3a and Extended Data Fig. 7a–d). Furthermore, almost all embryonic day (E) 14.5 fetal liver LSK CD150^{+} CD48^{-} HSCs were vWF^{+} (Fig. 3b), whereas in bone marrow, vWF^{-} HSCs were fully segregated from vWF^{+} HSCs 5 weeks after birth (Fig. 3c). Together, these observations indicated that vWF^{-} HSCs arise from vWF^{+} HSCs, and therefore that vWF^{+} platelet-primed HSCs developmentally and hierarchically precede vWF^{-} lymphoid-biased HSCs. Also at an earlier time point (7 weeks) after transplantation, vWF^{-} HSCs regenerated virtually only vWF^{-} LSK CD150^{+} CD48^{-} cells (Fig. 3d and Extended Data Fig. 10a). To test further the hierarchical relationship between vWF^{+} and vWF^{-} HSCs, we purified donor-derived vWF–eGFP^{-} and vWF–eGFP^{+} LSK CD150^{+} CD48^{-} cells from mice transplanted with vWF^{+} HSCs, and investigated their lineage reconstitution potentials after re-transplantation (Fig. 3d, e and Extended Data Fig. 10a). Whereas vWF–eGFP^{-} LSK CD150^{+} CD48^{-} cells generated from vWF–eGFP^{+} HSCs gave lymphoid-biased reconstitution, vWF–eGFP^{+} LSK CD150^{+} CD48^{-} cells reconstituted in a platelet-biased fashion (Fig. 3e). We also performed transplantations, omitting potentially biased HSC antigens (Extended Data Fig. 10b), of Lin^{-} vWF–eGFP^{+} and Lin^{-} vWF–eGFP^{-} cells generated from vWF^{+} and vWF^{-} HSCs, into secondary recipients. Potent platelet-biased reconstitution was exclusively associated with vWF–eGFP^{+} cells derived from vWF^{+} HSCs (Extended Data Fig. 10c). Low and inconsistent reconstitution of all blood cell lineages was observed from vWF–eGFP^{+} cells generated from vWF^{-} HSCs, probably reflecting the inability of vWF^{-} HSCs to regenerate vWF^{+} HSCs. Serially transplanted vWF^{+} cells derived from vWF^{+} HSCs also generated high levels of myeloid cells, but low levels of B and T cells, compatible with the vWF^{+} fraction also

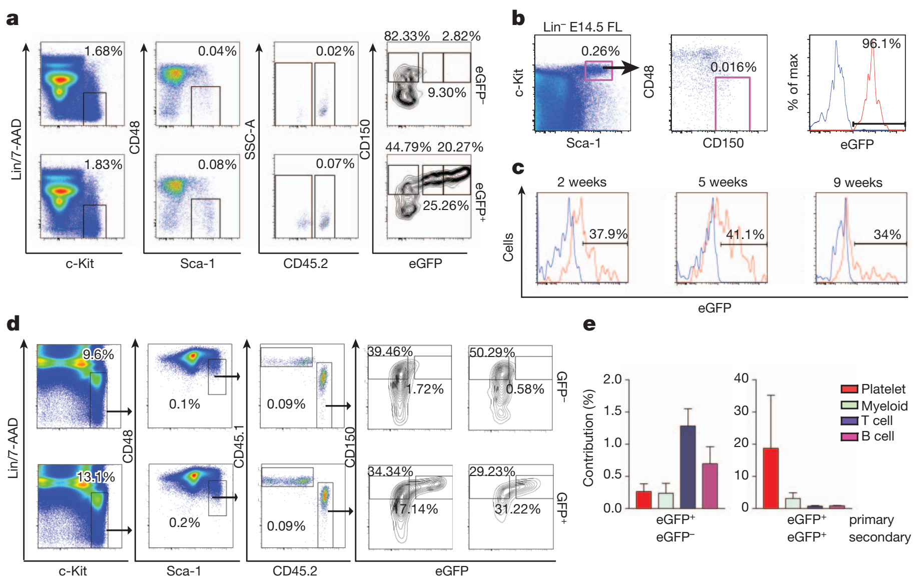


Figure 3 | Hierarchical relationship between vWF^{+} and vWF^{-} HSCs. **a**, HSC compartment 32 weeks after competitive transplantation of 50 vWF–eGFP^{+} or 50 vWF–eGFP^{-} LSK CD150^{+} CD48^{-} CD34^{-} HSCs. The first three plots from left show mean percentages of total bone marrow cells ($n = 13$ eGFP^{-} and 14 eGFP^{+}; 3 experiments). Right plots show mean frequencies of CD45.2^{+} LSK CD48^{-} CD150^{+} vWF–eGFP^{-}, CD45.2^{+} LSK CD48^{-} CD150^{+} vWF–eGFP^{low} and CD45.2^{+} LSK CD48^{-} CD150^{+} vWF-eGFP^{high} HSCs of total CD45.2^{+} LSK CD48^{-} cells. 7-AAD, 7-aminoactinomycin D. **b**, **c**, Mean vWF–eGFP expression in (**b**) E14.5 fetal liver (FL) LSK CD150^{+} CD48^{-} HSCs ($n = 15$), and (**c**) bone marrow LSK CD150^{+} CD48^{-} HSCs at 2 weeks ($n = 3$), 5 weeks ($n = 3$) and 9 weeks ($n = 2$) after birth. **d**, HSC compartment 7 weeks after competitive transplantation of vWF^{+} or vWF^{-} HSCs. Representative three left FACS plots show percentages relative to total bone marrow cells. Representative two right contour plots show frequencies of CD150^{+} vWF–eGFP^{-} and CD150^{+} vWF–eGFP^{+} HSCs relative to total CD45.2^{+} LSK CD48^{-} cells. **e**, Mean (± s.e.m.) peripheral blood test-cell lineage contribution 16 weeks after secondary transplantation of re-purified CD45.2^{+} eGFP^{-} and eGFP^{+} HSCs ($n = 4$). eGFP status of primary and secondary transplanted HSCs is indicated.

containing platelet/myeloid- and potentially myeloid-biased HSCs. Conversely, recipients of vWF$^-$ bone marrow cells, whether derived from vWF$^+$ or vWF$^-$ HSCs, showed a distinct lymphoid-biased reconstitution (Extended Data Fig. 10c), reflecting enhanced lymphopoiesis as confirmed by sustained regeneration of T-cell progenitors in the thymus (Extended Data Fig. 10d, e) and B-cell progenitors in bone marrow (Extended Data Fig. 10f, g). From these results, we conclude that vWF$^+$ platelet-primed HSCs give rise to vWF$^-$ lymphoid-biased HSCs.

Our results identify vWF$^+$ platelet-primed HSCs as a primitive, THPO-dependent HSC population capable of short-term (as well as stable long-term) platelet-biased reconstitution. Our studies not only establish the hierarchical relationship of vWF$^+$ platelet-primed HSCs at the apex of the HSC hierarchy and downstream vWF$^-$ lymphoid-biased HSCs lacking platelet-priming, they also shed new light on previously identified myeloid-biased HSCs[2–4,16]. Although in the short term vWF$^+$ platelet-primed HSCs seem to act more exclusively as platelet-biased HSCs, our finding that in the long term they contain stable platelet-biased as well as platelet/myeloid-biased HSCs, in both cases producing proportionally more myeloid than lymphoid cells, suggest that previous studies in the absence of methods to evaluate platelet output might have identified both platelet-biased and platelet/myeloid-biased HSCs as myeloid-biased stem cells[2–4,16]. The identification here of potent and stably platelet-biased HSCs capable also of short-term platelet-biased reconstitution provides an important therapeutic target for enhancing platelet reconstitution after chemotherapy and bone marrow transplantation.

METHODS SUMMARY

Mice. *Vwf-eGFP* BAC mice (CD45.2) were generated as described in the Methods.
Flow cytometry. vWF–eGFP$^+$ and vWF–eGFP$^-$ HSCs were analysed or sorted by FACS.
Microarray and qPCR analysis. Global gene expression profiling was performed using Affymetrix GeneChip arrays (MOE430 2.0) and data were analysed as described in Methods. qPCR expression analysis was performed with BioMark 48×48 Dynamic Array (Fluidigm).
***In vivo* transplantation studies.** vWF–eGFP$^+$ and vWF–eGFP$^-$ HSCs (CD45.2 allotype) were intravenously injected, along with wild-type unfractionated bone marrow cells (CD45.1), into lethally irradiated CD45.1 recipients (900 cGy, split dose). Secondary transplantation was performed into lethally irradiated CD45.1 recipients. For hierarchical experiments, vWF$^+$ and vWF$^-$ CD45.2$^+$ cells were re-sorted 6–7 weeks after primary transplantation and transplanted into lethally irradiated CD45.1 recipients. Peripheral blood and bone marrow reconstitution was assessed at the indicated time points by analysis of donor contribution to the myeloid, B-cell, T-cell and platelet lineages.
Statistical methods. Unless otherwise stated, data are presented as mean ± s.e.m. values. Two-tailed Student's *t*-tests were used to assess statistical significance for pairwise comparisons and one-way analysis of variance (ANOVA) for multiple comparisons.

Online Content Any additional Methods, Extended Data display items and Supplementary Information are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 August 2012; accepted 22 July 2013.
Published online 11 August 2013.

1. Nash, R. A., Gooley, T., Davis, C. & Appelbaum, F. R. The problem of thrombocytopenia after hematopoietic stem cell transplantation. *Oncologist* **1,** 371–380 (1996).
2. Müller-Sieburg, C. E., Cho, R. H., Thoman, M., Adkins, B. & Sieburg, H. B. Deterministic regulation of hematopoietic stem cell self-renewal and differentiation. *Blood* **100,** 1302–1309 (2002).
3. Dykstra, B. *et al.* Long-term propagation of distinct hematopoietic differentiation programs *in vivo*. *Cell Stem Cell* **1,** 218–229 (2007).
4. Copley, M. R., Beer, P. A. & Eaves, C. J. Hematopoietic stem cell heterogeneity takes center stage. *Cell Stem Cell* **10,** 690–697 (2012).
5. Akashi, K. *et al.* Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. *Blood* **101,** 383–389 (2003).
6. Adolfsson, J. *et al.* Identification of Flt3$^+$ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell* **121,** 295–306 (2005).
7. Månsson, R. *et al.* Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. *Immunity* **26,** 407–419 (2007).
8. Huang, H. & Cantor, A. B. Common features of megakaryocytes and hematopoietic stem cells: what's the connection? *J. Cell. Biochem.* **107,** 857–864 (2009).
9. de Sauvage, F. J. *et al.* Physiological regulation of early and late stages of megakaryocytopoiesis by thrombopoietin. *J. Exp. Med.* **183,** 651–656 (1996).
10. Forsberg, E. C., Bhattacharya, D. & Weissman, I. L. Hematopoietic stem cells: expression profiling and beyond. *Stem Cell Rev.* **2,** 23–30 (2006).
11. Denis, C. *et al.* A mouse model of severe von Willebrand disease: defects in hemostasis and thrombosis. *Proc. Natl Acad. Sci. USA* **95,** 9524–9529 (1998).
12. Kiel, M. J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121,** 1109–1121 (2005).
13. Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273,** 242–245 (1996).
14. Benveniste, P. *et al.* Intermediate-term hematopoietic stem cells with extended but time-limited reconstitution potential. *Cell Stem Cell* **6,** 48–58 (2010).
15. Kent, D. G. *et al.* Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential. *Blood* **113,** 6342–6350 (2009).
16. Morita, Y., Ema, H. & Nakauchi, H. Heterogeneity and hierarchy within the most primitive hematopoietic stem cell compartment. *J. Exp. Med.* **207,** 1173–1182 (2010).
17. Gekas, C. & Graf, T. CD41 expression marks myeloid-biased adult hematopoietic stem cells and increases with age. *Blood* **121,** 4463–4472 (2013).
18. Shimazu, T. *et al.* CD86 is expressed on murine hematopoietic stem cells and denotes lymphopoietic potential. *Blood* **119,** 4889–4897 (2012).
19. Goodell, M. A., Brose, K., Paradis, G., Conner, A. S. & Mulligan, R. C. Isolation and functional properties of murine hematopoietic stem cells that are replicating *in vivo*. *J. Exp. Med.* **183,** 1797–1806 (1996).
20. Pronk, C. J. *et al.* Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. *Cell Stem Cell* **1,** 428–442 (2007).
21. Yoshihara, H. *et al.* Thrombopoietin/MPL signaling regulates hematopoietic stem cell quiescence and interaction with the osteoblastic niche. *Cell Stem Cell* **1,** 685–697 (2007).
22. Qian, H. *et al.* Critical role of thrombopoietin in maintaining adult quiescent hematopoietic stem cells. *Cell Stem Cell* **1,** 671–684 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank the Biomedical Services at Oxford University, B. Wu and H. Ferry for technical assistance, and M. José Fernandez Nestosa, S. Luc and N. Buza-Vidas for discussions and experimental contributions. This work was supported by the Medical Research Council, UK Grant H4RPLK0 (to S.E.W.J.) and an MRC Program Grant and Strategic Award (to C.N.), by the European Commission FP7 CardioCell (to C.N.) and EuroSyStem (to S.E.W.J.) projects, and by the Association for International Cancer Research (to C.N.). A.S.-P. was supported by a postdoctoral BEOI fellowship from the Spanish Ministry of Education and Science.

Author Contributions S.E.W.J., C.N. and A.S.-P. designed the project, analysed data and wrote the manuscript. A.S.-P., I.C.M. and C.T.J. organized, performed and analysed most experiments. P.S.W., T.C.L. and A.M. participated in experimental design, statistical analysis and performed/analysed experiments. S.Mo. and S.T. performed analysis of gene expression data. C.C. contributed with cloning of targeting construct for the BAC transgenic mouse model. T.B.J. helped with peripheral blood reconstitution analysis in transplantation studies. O.C. performed stem-cell transplantation experiments. M.L. performed *in vitro* megakaryocyte culture experiments and helped with transplantations. S.Ma. performed PCR analysis of endothelial and osteoblastic cells. L.S., J.C.A.G., R.F., H.B., A.Gr., A.Ga., J.C. and P.T. were involved in reconstitution analysis. D.A. helped with single-cell qPCR experiments. S.-A.C. performed cell sorting. All authors read and approved the final manuscript.

Author Information The microarray data can be found in the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress/) with the accession number E-MEXP-3935. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.E.W.J. (sten.jacobsen@imm.ox.ac.uk) or C.N. (claus.nerlov@imm.ox.ac.uk).

LETTER

doi:10.1038/nature12616

Temperature triggers immune evasion by *Neisseria meningitidis*

Edmund Loh[1]*, Elisabeth Kugelberg[2]*, Alexander Tracy[1], Qian Zhang[2], Bridget Gollan[2], Helen Ewles[2], Ronald Chalmers[3], Vladimir Pelicic[2] & Christoph M. Tang[1,2]

Neisseria meningitidis **has several strategies to evade complement-mediated killing, and these contribute to its ability to cause septicaemic disease and meningitis. However, the meningococcus is primarily an obligate commensal of the human nasopharynx, and it is unclear why the bacterium has evolved exquisite mechanisms to avoid host immunity. Here we demonstrate that mechanisms of meningococcal immune evasion and resistance against complement increase in response to an increase in ambient temperature. We have identified three independent RNA thermosensors located in the 5′ untranslated regions of genes necessary for capsule biosynthesis, the expression of factor H binding protein, and sialylation of lipopolysaccharide, which are essential for meningococcal resistance against immune killing[1,2]. Therefore increased temperature (which occurs during inflammation) acts as a 'danger signal' for the meningococcus, enhancing its defence against human immune killing. Infection with viral pathogens, such as influenza, leads to inflammation in the nasopharynx with an increased temperature and recruitment of immune effectors[3,4]. Thermoregulation of immune defence could offer an adaptive advantage to the meningococcus during co-infection with other pathogens, and promote the emergence of virulence in an otherwise commensal bacterium.**

Neisseria meningitidis is an obligate human pathogen and important cause of sepsis and meningitis[5], with peaks of disease often preceded by influenza outbreaks in temperate climates[6]. The bacterium has evolved exquisite mechanisms to evade immune responses[7], including expression of a polysaccharide capsule (containing sialic acid in serogroup B, C, Y and W strains)[8,9], sialylation of lipopolysaccharide, and recruitment of the human complement regulator factor H, via high-affinity interactions with bacterial factor H binding protein[1,2]. The reasons why such mechanisms have evolved in an otherwise commensal bacterium is uncertain as systemic infection represents an evolutionary dead end.

Previously we found that resistance of the meningococcus against complement-mediated killing is enhanced after insertion of the mobile element IS*1301* into the 134-base-pair (bp) intergenic region (IGR) between the *css* (encoding capsule biosynthesis) and *ctr* (capsule export) operons in the capsule biosynthesis locus (*cps*)[10,11]. To identify other changes modulating serum resistance, we subjected the *N. meningitidis* strain S3 (ref. 11) to serial passage in 6% human serum; within six rounds, S3 became as resistant to complement-mediated killing as a strain with IS*1301* in the IGR (Supplementary Fig. 1). We further characterized six passaged strains (selected serum resistant, SSR1–6) that were more resistant to complement than S3 (Fig. 1a). Resistance did not result from insertion of IS*1301*; instead five strains (all except SSR2) had lost a single copy of a duplicated 8-nucleotide sequence (TATACTTA) located 15 nucleotides upstream of the CssA start codon in the 5′ untranslated region (5′ UTR) of *css* messenger RNA (Fig. 1b, c and Supplementary Fig. 2) and have increased levels of CssA, which catalyses an early step in capsule biosynthesis[12]. Comparison of S3 with SSR, and isogenic strains (both containing an antibiotic-resistance cassette downstream of the *css* operon) with one (Δ8) or two (wild-type) copies of the 8-bp sequence demonstrates that loss of 8 bp causes increased CssA levels and capsule expression (Fig. 1d, e). The increased serum resistance of SSR2 resulted from reduced levels of PorA (Fig. 1b), a target of bactericidal antibodies[13].

To define how the 8-bp sequence contributes to capsule expression, we performed DNA foot-printing of the IGR with integration host factor (IHF) and factor for inversion stimulation (FIS). Despite potential recognition sites for these proteins[10,14], there was no difference in IHF or FIS binding to the wild-type and Δ8 IGRs (Supplementary Fig. 3). Analysis of translational reporters in the *N. meningitidis cps* demonstrated that loss of 8 bp leads to a marked increase in reporter activity of Css but not Ctr (Supplementary Figs 4 and 5), demonstrating that Δ8 influences the capsule biosynthesis operon. Furthermore, significantly increased Css reporter activity with the Δ8 IGR was evident in 38 transcription factor mutants (Supplementary Table 1 and Supplementary Fig. 5), and northern blot analysis demonstrates that *css* mRNA levels are unaffected by the number of copies of the 8 bp (Fig. 1f). Therefore, the 8-bp sequence affects CssA post-transcriptionally.

We noticed that the *css* 5′ UTR is predicted to form a stem–loop structure that includes the ribosome-binding site (RBS) (Fig. 2a), consistent with an RNA thermosensor[15]. In RNA thermosensors, the transcript assumes a hairpin structure at lower temperatures that occludes the RBS, and stalls protein translation; higher temperatures destabilize the secondary structure which allows translation. The Δ8 mRNA, on the other hand, is predicted to form a limited stem–loop with a single-stranded region by the RBS (Fig. 2a). Consistent with a thermosensor, CssA levels increase in *N. meningitidis* grown at increasing temperatures (Fig. 2b). By contrast, loss of 8 bp leads to increased CssA levels at lower temperatures and less pronounced increase at higher temperatures, suggesting that this change disrupts the thermosensor and dysregulates capsule biosynthesis.

RNA thermosensors should function in a heterologous host. Similar to *N. meningitidis*, thermal regulation of CssA was evident in *Escherichia coli* containing CssA and the wild-type IGR on a plasmid; CssA expression was dysregulated with the Δ8 IGR (Fig. 2c). Additionally, *in vitro* transcription/translation assays demonstrated that CssA synthesis increased with an increase in temperature and on loss of one copy of the 8-bp sequence (Fig. 2d) in the absence of any transcription factor. Furthermore, we introduced nucleotide changes into the 5′ UTR predicted to alter the stability of the thermosensor, including substitutions at the same position ($+92_{U/C}$ or $+92_{U/G}$; Fig. 2a) expected to have opposing effects. Expression of CssA from plasmids containing these changes was consistent with a thermosensor in the 5′ UTR (Fig. 2e). We also performed RNA toe-printing at 30 °C, 37 °C and 42 °C to assess binding of ribosomes to the nascent *css* transcript. The results demonstrate that ribosome binding is enhanced to mRNA from the Δ8 compared to the wild-type 5′ UTR, with differences most marked at 30 °C (Supplementary Fig. 6). Finally we found that thermoregulation of CssA is

[1]Sir William Dunn School of Pathology, University of Oxford, Sir Parks Road, Oxford OX1 3RE, UK. [2]Centre for Molecular Microbiology and Infection, Imperial College London, London SW7 2AZ, UK. [3]School of Biomedical Sciences, University of Nottingham, Nottingham NG7 2NR, UK.
*These authors contributed equally to this work.

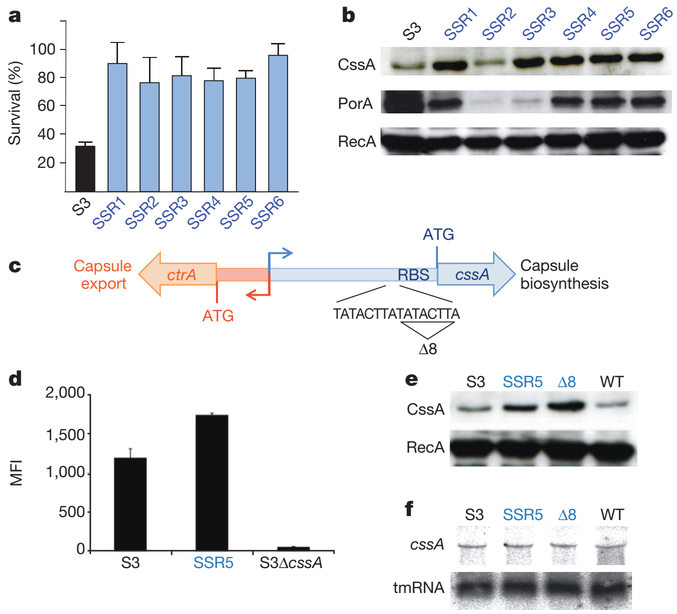


Figure 1 | Loss of 8 bp increases capsule expression and complement resistance. **a**, **b**, Selected serum resistant (SSR1–6) strains demonstrate enhanced resistance in human serum against complement-mediated killing (**a**), and increased expression of CssA (**b**, except SSR2; RecA loading control) compared with the parental strain, S3. **c**, The 134-bp *css*/*ctr* intergenic region showing the ribosome-binding site (RBS), transcriptional start sites (arrows) and open reading frames, and location of the duplicated 8-bp sequence. **d**, Resistant strains exhibit increased capsule expression by FACS. MFI, mean fluorescence index. **e**, CssA expression in isogenic strains with one or two copies of the 8-bp sequence. **f**, Northern blot analysis demonstrates that increased CssA levels are not associated with increased *cssA* mRNA. tmRNA, transfer-messenger RNA. Findings were confirmed in three biological replicates. Error bars show s.d. of experiments performed in triplicate.

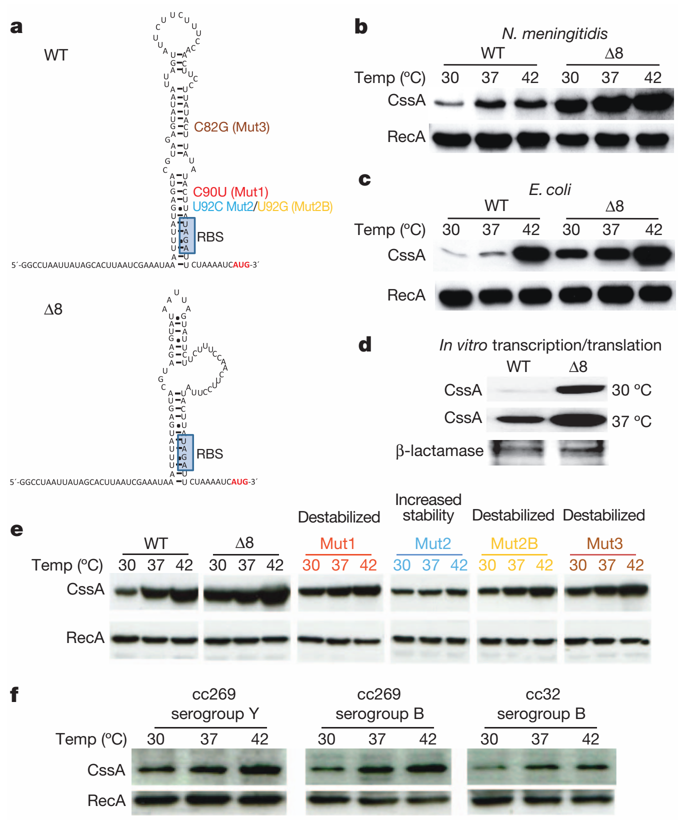


Figure 2 | Capsule gene expression is governed by an RNA thermosensor. **a**, Predicted secondary structure of the *css* 5′ UTR with two (wild type, WT) or one (Δ8) copy of the 8-nucleotide sequence; the RBS is indicated. **b**, Western blot analysis of CssA levels in isogenic *N. meningitidis* strains with a wild-type IGR or containing a single copy of the 8-bp sequence (Δ8) grown at different temperatures. **c**, **d**, Thermoregulation of CssA expression is detected in *E. coli* by western blot analysis (**c**), and by *in vitro* transcription/translation assays with purified *E. coli* RNA polymerase, with β-lactamase included as a control (**d**). **e**, Mutational analysis of the 5′ UTR. *E. coli* expressing CssA with modifications (shown in **a**) grown at temperatures indicated; the predicted effects shown. **f**, Thermoregulation of CssA is evident in *N. meningitidis* across different capsular serogroups and lineages. cc, clonal complex. Findings were confirmed with three biological replicates.

evident in strains of different capsular serogroups and hypervirulent lineages (Fig. 2f). Together, these results confirm the presence of an RNA thermosensor controlling the capsule biosynthesis operon in *N. meningitidis* across a range of strains.

Previous reports of RNA thermosensors in pathogens are restricted to facultative organisms and govern transcription factors that mediate transition of bacteria from external to internal environments[15–18]. As the meningococcus is an obligate commensal residing in the nasopharynx, we compared the dynamic response of the canonical PrfA *Listeria* thermosensor with the meningococcal Css thermosensor (which directs a single pathway). The *prfA* and *css* 5′ UTRs and their promoters were fused to green fluorescent protein (GFP) in the plasmid pEGFP-N2, and *in vitro* transcription/translation was performed at different temperatures (Fig. 3a). Protein synthesis regulated by the *prfA* thermosensor was barely detected up to 36 °C, but markedly increased at higher temperatures. In contrast to this on/off switching, the *css* thermosensor displayed a gradual increase over physiologically relevant temperatures (acting like a rheostat). The difference can be explained by the relatively high GC content and distribution of GC bonds flanking the RBS in the *prfA* 5′ UTR (Fig. 3b); in the bacteria, the *Listeria* thermosensor mediates changes over a marked temperature shift as the bacterium migrates from the external environment into its host, whereas *N. meningitidis* is in virtual constant contact with its host and therefore less exposed to wide temperature fluctuations.

Next we determined the prevalence of IGR polymorphisms in 265 meningococcal disease isolates. The wild-type sequence (two copies of the 8-nucleotide sequence) was most frequently found (201 of 265, 75.8%); of note, no polymorphism was detected in the 5′ UTR with two copies of the 8 nucleotides. Strains with a Δ8 accounted for the remainder (64 out of 265, 24.2%; Fig. 3c), although this deletion was accompanied by two substitutions (TAT<u>G</u>C<u>C</u>TAT; altered bases underlined, $\Delta8_{AT/GC}$) in most instances (50 out of 64 isolates, 78.1%). Only a few strains (8 out of 64, 12.5%) had the Δ8 sequence with no changes, and the single, TAT<u>G</u>CTTAT polymorphism ($\Delta8_{A/G}$) was present in the other strains (6 out of 64, 9.4%), whereas TATAC<u>C</u>TAT ($\Delta8_{T/C}$) was never detected. We analysed the effect of these polymorphisms in *E. coli* plasmid reporters (Fig. 3d), and found that whereas $\Delta8_{A/G}$ partially restores CssA thermal regulation after deletion of 8 nucleotides, $\Delta8_{AT/GC}$ re-establishes thermosensing (consistent with its predicted structure and RNA toe-printing; Supplementary Figs 6 and 7). By contrast, $\Delta8_{T/C}$ (which was never found) leads to markedly increased CssA expression at all temperatures. Of note, the compensatory polymorphisms occur at similar frequencies in different clonal complexes of the meningococcus, consistent with them arising on several occasions.

The conservation of the wild-type sequence and prevalence of compensatory polymorphisms emphasize the importance of capsule thermoregulation in the meningococcus. Hence, we examined whether other factors involved in immune escape are subject to similar regulation. Of note, expression of factor H binding protein (which recruits the host complement regulator factor H) and Lst (necessary for lipopolysaccharide sialylation[19]) also increase with increasing temperature (Fig. 3e), in contrast to proteins not involved in immune escape (such as PorB, Pilin, RmpM and RecA), which are unaffected by temperature (Fig. 3f). To define the mechanism of thermosensing of fHbp and Lst,

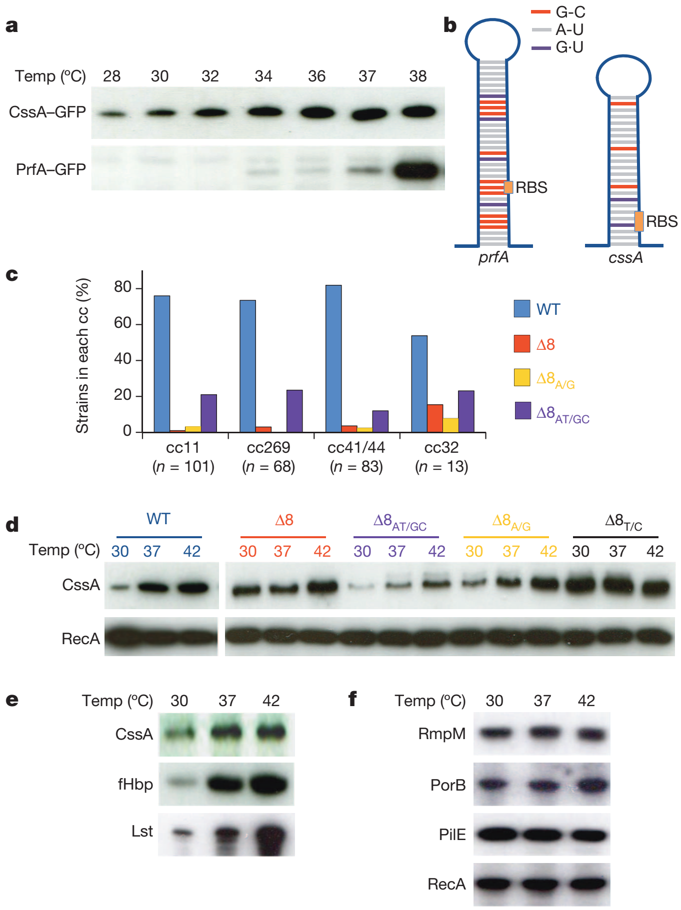


Figure 3 | Thermosensing of immune evasion mechanisms. **a**, **b**, *In vitro* transcription/translation assays (**a**) and predicted structure (**b**) of the Css and PrfA thermosensors modulating GFP expression in *E. coli* at a range of temperatures. **c**, Sequence analysis of the IGR in different hypervirulent lineages; clonal complexes shown. Loss of a copy of the 8-bp sequence (Δ8) is associated with a 2-bp substitution in the remaining single 8-bp sequence (that is, $\Delta8_{AT/GC}$), which is never detected in a strain with two copies of the 8-bp sequence. **d**, Effect of the polymorphisms examined in *E. coli*; the most abundant polymorphism $\Delta8_{AT/GC}$ restores thermosensing to wild-type levels in the presence of a single 8-bp sequence, whereas $\Delta8_{A/G}$ substitution partially restore thermosensing; $\Delta8_{T/C}$ leads to higher CssA levels than Δ8. **e**, **f**, Western blot analysis demonstrates that levels of proteins mediating immune evasion, fHbp and Lst, increase in *N. meningitidis* with an increase in temperature (**e**), whereas other proteins (indicated) are unaffected (**f**). Blots are representative of experiments performed on at least three occasions.

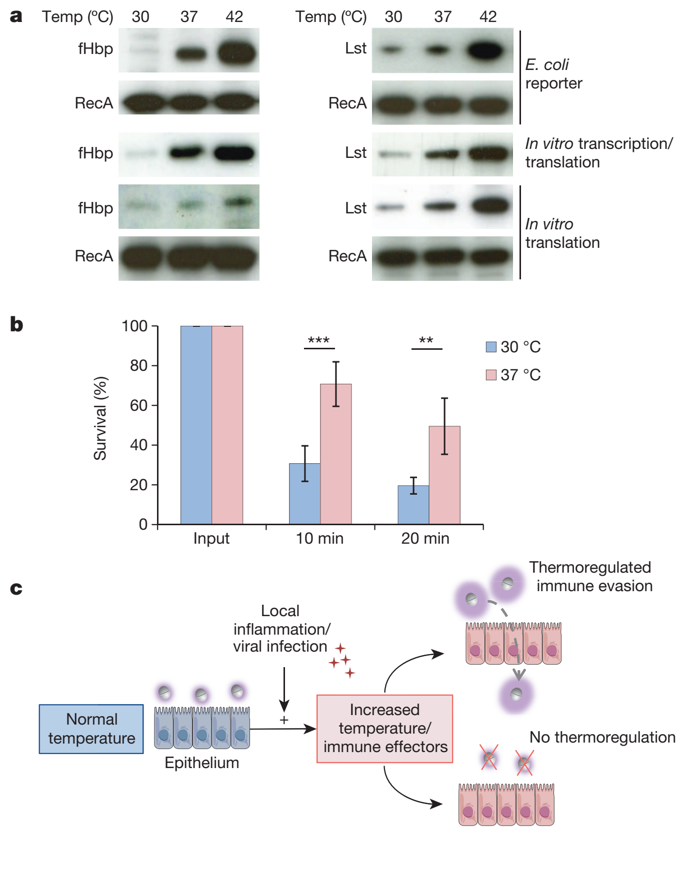


Figure 4 | Temperature influences complement escape by *N. meningitidis*. **a**, Thermoregulation of fHbp and Lst expression in *E. coli* reporters detected by western blot analysis (RecA, loading control), in *in vitro* transcription/translation assays, and in *in vitro* translation assays of fHbp and Lst (RecA, expression control). Blots are representative of experiments performed on at least three occasions. **b**, Bacteria incubated for 1 h at 37 °C are more resistant to complement-mediated killing than those at 30 °C: the assay was performed on four separate occasions, and error bars denote s.e.m. $^{**}P < 0.01$; $^{***}P < 0.001$ (Student's *t*-test). **c**, Inflammation at the epithelial surface in the nasopharynx (in response to viral co-infection) would increase local temperature and recruitment of innate immune effectors. Thermoregulation of microbial defence mechanisms would prevent bystander killing and enhance systemic dissemination to warmer body sites that are replete with immune effectors.

we analysed *E. coli* plasmids containing these genes and observed thermal regulation (Fig. 4a). Furthermore, thermoregulation of these proteins in *N. meningitidis* was independent of 38 transcription factors (Supplementary Fig. 8), and thermoregulation of fHbp and Lst was detected in both *in vitro* transcription/translation assays (with fixed amounts of DNA) and *in vitro* translation assays (with fixed amounts of RNA; Fig. 4a), indicating that the 5′ UTRs of *fHbp* and *lst* contain RNA thermosensors, consistent with secondary RNA structure predictions (Supplementary Fig. 9).

Temperature therefore acts as a danger signal for the meningococcus, prompting the bacterium to enhance expression mechanisms of immune evasion via three independent thermosensors dedicated to single proteins or pathways. To determine the influence of temperature on meningococcal complement resistance, bacteria were grown at 30 °C then incubated at 30 °C or 37 °C for one hour. Bacteria that had been equilibrated at the increased temperature expressed more CssA, fHbp and Lst (Supplementary Fig. 10) and were significantly more resistant against complement than those left at 30 °C (Fig. 4b), demonstrating that thermal regulation of immune defence mechanisms has a marked effect on bacterial survival in the presence of human complement.

An increase in temperature is a cardinal feature of inflammation, which is associated with extravasation of serum components and recruitment of phagocytes. In these circumstances, increasing expression of factors necessary for immune evasion would provide a considerable adaptive advantage to a microbe in the nasopharynx (Fig. 4c). Influenza provokes complement activation in the upper airway[3], a rise in core temperature, and elaborates neuraminidase that could damage the meningococcal capsule[20]. By sensing local inflammation, thermal regulation of immune defence by the meningococcus would allow adaption to changes in the nasopharynx during infection with influenza[4,21] and other respiratory pathogens; microbes unable to sense and protect themselves against inflammatory responses will face elimination from this habitat (Fig. 4c). Furthermore, thermoregulation by the meningococcus would provide an advantage to bacteria entering the bloodstream, which is at higher temperatures than the nasopharynx[22]. Therefore, traits that are beneficial for bacterial colonization during coinfection and inflammation (such as thermoregulation of immune defence) will, by unfortunate coincidence, promote the virulence of otherwise commensal microbes.

METHODS SUMMARY

N. meningitidis and *E. coli* were grown on brain heart infusion (BHI) broth (Oxoid) and Luria–Bertani (Oxoid) agar, respectively. For liquid growth, bacteria were grown in broth to an absorbance (A) at 600 nm of ~0.4. For human serum sensitivity assays, bacterial strains were grown on BHI agar plates overnight and re-suspended in PBS (Oxoid). Bacteria were diluted in DMEM-glutaMAX medium (Invitrogen), and incubated with different concentrations of normal human sera at 37 °C in presence of 5% CO_2 for 1 h. Protein samples for western blotting were separated on 12% polyacrylamide gels and transferred onto PVDF (polyvinylidene difluoride) membranes (Milipore), which were incubated with primary then secondary antibodies. β-galactosidase assays were performed as described previously using *o*-nitrophenyl-β-galactoside (Thermo Scientific Pierce) (4 mg ml^{-1}) as the substrate; reactions were stopped by the addition of $NaCO_3$. FACS analysis to determine the amount of capsule has been described previously[11]. To determine the effect of temperature on complement sensitivity, bacteria were grown in liquid media to mid-logarithmic phase at 30 °C, then split and incubated at 30 °C or 37 °C for a further 1 h. A total of 1×10^6 colony-forming units were incubated with serial dilutions of pooled human immune serum for either 10 or 20 min, and the proportion of bacteria surviving was determined; significant differences were examined with Student's *t*-test. Total RNA was isolated using the RNAeasy Miniprep kit (Qiagen). Northern blotting was performed on 1.2% agarose, the RNA transferred to Hybond-N$^+$ membranes (GE Healthcare) and hybridized in Rapid-hyb buffer (GE Healthcare). *In vitro* transcription/translation was performed with an *E. coli* S30 Extract system for Linear Templates *in vitro* transcription/translation kit (Promega) according to the manufacturer's instructions. RNA secondary structures were predicted using the Vienna RNA package (http://www.tbi.univie.ac.at/~ivo/RNA/).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 April; accepted 29 August 2013.
Published online 25 September 2013.

1. Schneider, M. C. *et al.* Functional significance of factor H binding to *Neisseria meningitidis*. *J. Immunol.* **176,** 7566–7575 (2006).
2. Schneider, M. C. *et al. Neisseria meningitidis* recruits factor H using protein mimicry of host carbohydrates. *Nature* **458,** 890–893 (2009).
3. Bjornson, A. B., Mellencamp, M. A. & Schiff, G. M. Complement is activated in the upper respiratory tract during influenza virus infection. *Am. Rev. Respir. Dis.* **143,** 1062–1066 (1991).
4. Gentile, D. *et al.* Increased interleukin-6 levels in nasal lavage samples following experimental influenza A virus infection. *Clin. Diagn. Lab. Immunol.* **5,** 604–608 (1998).
5. Stephens, D. S., Greenwood, B. & Brandtzaeg, P. Epidemic meningitis, meningococcaemia, and *Neisseria meningitidis*. *Lancet* **369,** 2196–2210 (2007).
6. Cartwright, K. A. *et al.* Influenza A and meningococcal disease. *Lancet* **338,** 554–557 (1991).
7. Schneider, M. C., Exley, R. M., Ram, S., Sim, R. B. & Tang, C. M. Interactions between *Neisseria meningitidis* and the complement system. *Trends Microbiol.* **15,** 233–240 (2007).
8. Stephens, D. S., Spellman, P. A. & Swartley, J. S. Effect of the (α2→8)-linked polysialic acid capsule on adherence of *Neisseria meningitidis* to human mucosal cells. *J. Infect. Dis.* **167,** 475–478 (1993).
9. Hammerschmidt, S. *et al.* Contribution of genes from the capsule gene complex (cps) to lipooligosaccharide biosynthesis and serum resistance in *Neisseria meningitidis*. *Mol. Microbiol.* **11,** 885–896 (1994).
10. Tzeng, Y. L. *et al.* Transcriptional regulation of divergent capsule biosynthesis and transport operon promoters in serogroup B *Neisseria meningitidis*. *Infect. Immun.* **69,** 2502–2511 (2001).
11. Uria, M. J. *et al.* A generic mechanism in *Neisseria meningitidis* for enhanced resistance against bactericidal antibodies. *J. Exp. Med.* **205,** 1423–1434 (2008).
12. Ganguli, S. *et al.* Molecular cloning and analysis of genes for sialic acid synthesis in *Neisseria meningitidis* group B and purification of the meningococcal CMP-NeuNAc synthetase enzyme. *J. Bacteriol.* **176,** 4583–4589 (1994).
13. van der Ende, A., Hopman, C. T. & Dankert, J. Multiple mechanisms of phase variation of PorA in *Neisseria meningitidis*. *Infect. Immun.* **68,** 6685–6690 (2000).
14. von Loewenich, F. D., Wintermeyer, E., Dumig, M. & Frosch, M. Analysis of transcriptional control mechanisms of capsule expression in *Neisseria meningitidis*. *Int. J. Med. Microbiol.* **291,** 361–369 (2001).
15. Johansson, J. *et al.* An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes*. *Cell* **110,** 551–561 (2002).
16. Hoe, N. P. & Goguen, J. D. Temperature sensing in *Yersinia pestis*: translation of the LcrF activator protein is thermally regulated. *J. Bacteriol.* **175,** 7901–7909 (1993).
17. Morita, M., Kanemori, M., Yanagi, H. & Yura, T. Heat-induced synthesis of σ^{32} in *Escherichia coli*: structural and functional dissection of *rpoH* mRNA secondary structure. *J. Bacteriol.* **181,** 401–410 (1999).
18. Böhme, K. *et al.* Concerted actions of a thermo-labile regulator and a unique intergenic RNA thermosensor control *Yersinia* virulence. *PLoS Pathog.* **8,** e1002518 (2012).
19. Kahler, C. M. *et al.* The (α2→8)-linked polysialic acid capsule and lipooligosaccharide structure both contribute to the ability of serogroup B *Neisseria meningitidis* to resist the bactericidal activity of normal human serum. *Infect. Immun.* **66,** 5939–5947 (1998).
20. Rameix-Welti, M. A. *et al.* Influenza A virus neuraminidase enhances meningococcal adhesion to epithelial cells through interaction with sialic acid-containing meningococcal capsules. *Infect. Immun.* **77,** 3588–3595 (2009).
21. Lijek, R. S. & Weiser, J. N. Co-infection subverts mucosal immunity in the upper respiratory tract. *Curr. Opin. Immunol.* **24,** 417–423 (2012).
22. Cole, P. Recordings of respiratory air temperature. *J. Laryngol. Otol.* **68,** 295–307 (1954).

Supplementary Information is available in the online version of the paper.

Acknowledgements Work in C.M.T.'s laboratory is funded by Programme Grants from the Wellcome Trust and the Medical Research Council, and by the Oxford Martin School. E.L. was supported by an International Postdoctoral fellowship from the Swedish Research Council (Dnr: 524-2010-6726) and E.K. was supported by an EMBO long-term fellowship. We are grateful to J. Johansson, S. Busby and other members of the Tang group for discussions, and to B. Kallipolitis for providing *E. coli* 30S ribosomes.

Author Contributions E.L., E.K., A.T., Q.Z., B.G. and H.E. performed the experiments and analysed the data. R.C. and V.P. oversaw the DNA foot-printing and construction of transcription factor mutants. C.M.T. provided overall direction, and wrote the manuscript with input from E.L., E.K. and R.C., and the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.M.T. (christoph.tang@path.ox.ac.uk).

doi:10.1038/nature12515

The genesis and source of the H7N9 influenza viruses causing human infections in China

Tommy Tsan-Yuk Lam[1,2,3]*, Jia Wang[1,3]*, Yongyi Shen[1,3,4]*, Boping Zhou[2], Lian Duan[2,3], Chung-Lam Cheung[3], Chi Ma[1,3], Samantha J. Lycett[5], Connie Yin-Hung Leung[3], Xinchun Chen[2], Lifeng Li[1,2,3], Wenshan Hong[1], Yujuan Chai[2,3], Linlin Zhou[3], Huyi Liang[1,2,3], Zhihua Ou[1,2,3], Yongmei Liu[1,3], Amber Farooqui[6], David J. Kelvin[6], Leo L. M. Poon[2,3], David K. Smith[1,3], Oliver G. Pybus[7,8], Gabriel M. Leung[1,3], Yuelong Shu[9], Robert G. Webster[10], Richard J. Webby[10], Joseph S. M. Peiris[2,3], Andrew Rambaut[5,11], Huachen Zhu[1,2,3] & Yi Guan[1,2,3]

A novel H7N9 influenza A virus first detected in March 2013 has since caused more than 130 human infections in China, resulting in 40 deaths[1,2]. Preliminary analyses suggest that the virus is a reassortant of H7, N9 and H9N2 avian influenza viruses, and carries some amino acids associated with mammalian receptor binding, raising concerns of a new pandemic[1,3,4]. However, neither the source populations of the H7N9 outbreak lineage nor the conditions for its genesis are fully known[5]. Using a combination of active surveillance, screening of virus archives, and evolutionary analyses, here we show that H7 viruses probably transferred from domestic duck to chicken populations in China on at least two independent occasions. We show that the H7 viruses subsequently reassorted with enzootic H9N2 viruses to generate the H7N9 outbreak lineage, and a related previously unrecognized H7N7 lineage. The H7N9 outbreak lineage has spread over a large geographic region and is prevalent in chickens at live poultry markets, which are thought to be the immediate source of human infections. Whether the H7N9 outbreak lineage has, or will, become enzootic in China and neighbouring regions requires further investigation. The discovery here of a related H7N7 influenza virus in chickens that has the ability to infect mammals experimentally, suggests that H7 viruses may pose threats beyond the current outbreak. The continuing prevalence of H7 viruses in poultry could lead to the generation of highly pathogenic variants and further sporadic human infections, with a continued risk of the virus acquiring human-to-human transmissibility.

After the initial reports of H7N9 influenza infection in humans, field surveillance was conducted during 4–18 April 2013 in Wenzhou (Zhejiang province, 500 km south of Shanghai) and Rizhao (Shandong province, 600 km north of Shanghai), which both border the main outbreak region, and in Shenzhen (Guangdong province, 1,200 km south of Shanghai), an area that has not reported human cases (Supplementary Fig. 1). A total of 1,341 pairs of oropharyngeal and cloacal samples were collected from chickens, ducks, geese, pigeons, partridges and quail. A further 1,006 faecal and water samples from live poultry markets (LPMs), farms and wetlands were also collected (Supplementary Table 1). A total of 388 haemagglutin-positive agents were isolated (10.5% of samples), of which 60 and 85 represented H7 and H9 influenza A viruses. The remaining positive isolates represented other subtypes of influenza A virus or avian paramyxovirus (Supplementary Table 1).

H7 influenza A viruses were only detected in Wenzhou and Rizhao, and only in LPMs. All H7 isolates from Rizhao were H7N9 viruses, whereas those from Wenzhou were all H7N7 viruses, except for two duck isolates that were H7N2 and H7N3 viruses. All H9 isolates were H9N2 viruses (80 from LPMs, 5 from farms). At LPMs in Wenzhou, the H7 virus was at its highest prevalence in chickens (10.1%; 46 out of 457), followed by ducks (2.4%; 3 out of 125) and pigeons (1.6%; 3 out of 188). In Rizhao, LPM H7N9 viruses were only found in chickens (0.7%; 8 out of 1,113). Of the chicken isolates, 100% of H7N9, 65.3% of H7N7 and 94.8% of H9N2 viruses were from oropharyngeal swabs (Supplementary Table 1), suggesting that these H7N9 and H7N7 viruses might replicate in the upper respiratory tract of terrestrial poultry, similar to the enzootic H9N2 viruses[6].

These samples were sequenced to investigate the evolutionary history of avian influenza viruses implicated in the current outbreak of H7N9 infections of humans and poultry. Full genome sequences were obtained for 34 H7N7, 4 H7N9 and 19 H9N2 isolates. The H7 and N7/9 genes of 16 mixed H7/H9 infections were sequenced (Supplementary Table 1), as were 3 H7N9 and 3 H7N7 samples that had multiple H9N2-like internal gene segments. The H7 haemagglutinin gene sequences of the H7N9 viruses isolated from chickens in Rizhao formed a tight monophyletic group (Fig. 1a, lineage 'b') with previously reported human and avian viruses from the current H7N9 outbreak. This was most closely related to a group comprising mainly H7N7 viruses obtained from Wenzhou chickens, ducks and pigeons (Fig. 1a, lineage 'c'). All viruses isolated from chickens in these two groups had internal gene complexes that were closely related to those present in co-circulating H9N2 viruses.

To examine the genesis of these H7N9 and H7N7 viruses, we sequenced 197 archived isolates of H7, N9, N7 and H9N2 viruses, obtained during previous influenza surveillance between 2000 and 2013 in southern China (Supplementary Fig. 1). These sequences were analysed together with those obtained in our post-outbreak surveillance, plus all closely related sequences from public databases (see Methods). H7 influenza viruses from East Asian migratory waterfowl were introduced into domestic ducks in China on several occasions during the past decade (Fig. 1a and Supplementary Fig. 2). In 2009–2010, H7 viruses with five NA subtypes were found in duck farms and LPMs in Jiangxi, suggesting an epidemiological bridge from migratory birds to sentinel farm ducks and then to market birds[7] (Supplementary Fig. 2). These introductions to domestic birds persisted for less than two years, except for the introduction of H7N3 viruses (initially isolated in Fujian and Zhejiang in 2010–2011) that led to the 2013 H7N9 outbreak lineage viruses and the H7N7 viruses from Wenzhou (Fig. 1a, lineage 'a').

Previous analyses have suggested that the N9 gene of the H7N9 outbreak lineage was derived from wild bird viruses in Europe and Korea[3,4].

[1]Joint Influenza Research Centre (SUMC/HKU), Shantou University Medical College, Shantou 515041, China. [2]State Key Laboratory of Emerging Infectious Diseases (HKU-Shenzhen Branch), Shenzhen Third People's Hospital, Shenzhen 518112, China. [3]State Key Laboratory of Emerging Infectious Diseases/Centre of Influenza Research, School of Public Health, The University of Hong Kong, Hong Kong SAR, China. [4]State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen 361102, China. [5]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK. [6]Joint Vaccine Research Centre (SUMC/UHN), Shantou University Medical College, Shantou 515041, China. [7]Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. [8]Metabiota, San Francisco, California 94104, USA. [9]National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Key Laboratory for Medical Virology, National Health and Family Planning Commission, Beijing 102206, China. [10]Division of Virology, Department of Infectious Diseases, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [11]Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892-2220, USA.
*These authors contributed equally to this work.

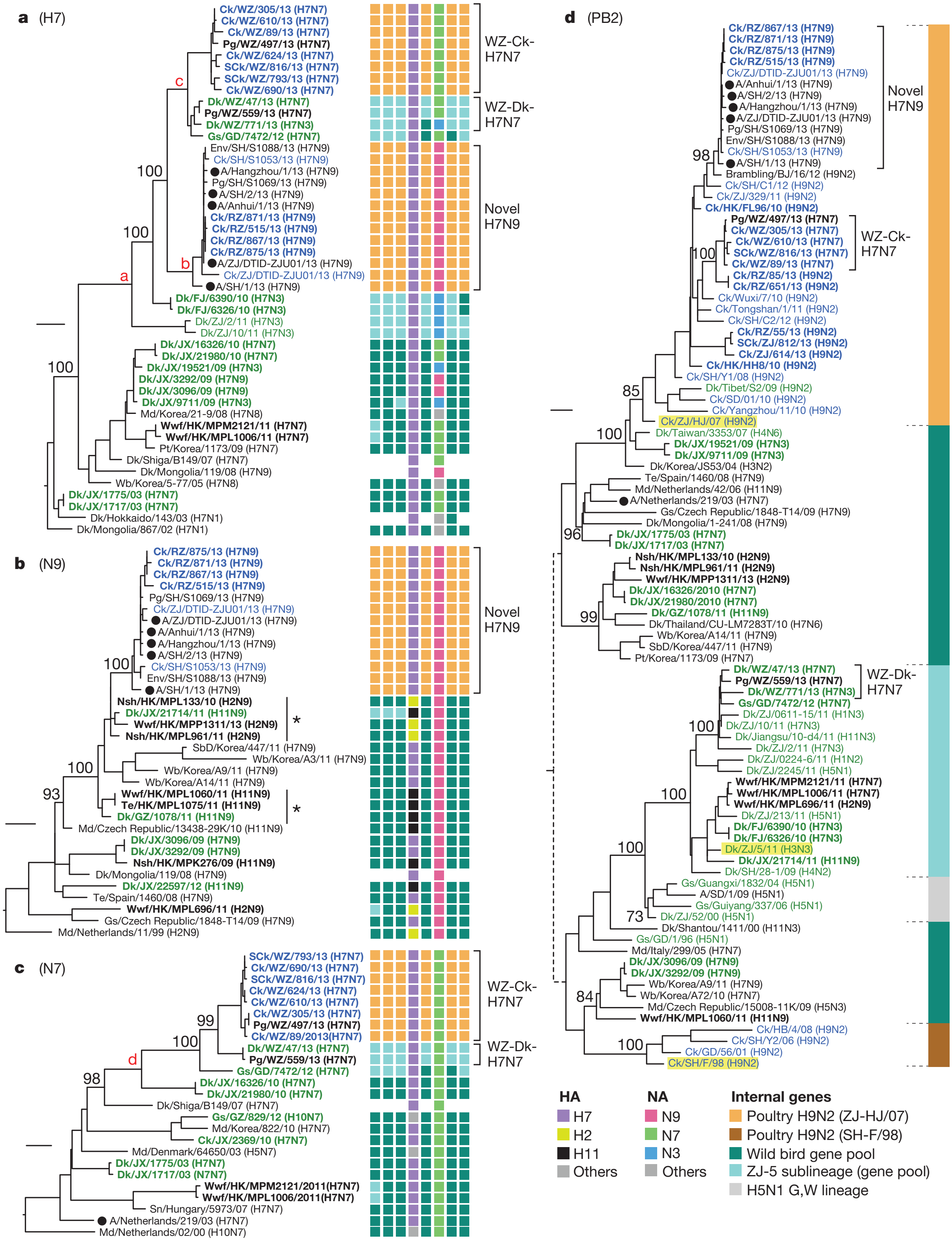


Figure 1 | Phylogenies of haemagglutinin, neuraminidase and PB2 genes. **a–d**, Phylogenies of H7 haemagglutinin ($n = 46$) (**a**), N9 neuraminidase ($n = 34$) (**b**), N7 neuraminidase ($n = 25$) (**c**) and PB2 ($n = 93$) (**d**) genes. Sequences reported in this study have their taxa names shown in bold. Genotypes of the influenza viruses are shown on the right (**a**–**c**) as eight coloured blocks representing each gene segment (from left to right: PB2, PB1, polymerase acidic, haemagglutinin, nucleoprotein, neuraminidase, matrix and non-structural; absent if the sequence is unavailable) with the colour indicating the subtype (for haemagglutinin, neuraminidase) or lineage (internal genes; indicated by the solid vertical line in **d**) of that segment. Bootstrap support values (%) from 1,000 pseudo-replicates are shown for selected lineages. Support values for lineages 'a'–'d' were all 100%. The scale bar to the left of each tree represents 0.01 substitutions per site. Asterisks in **b** denote N9 sub-lineages linking the viruses of domestic ducks and wild birds. Host species are: Ck (chicken), Dk (duck), Gs (goose), Md (mallard), Nsh (Northern shoveler), Pg (pigeon), Pt (pintail), SCk (silkie chicken), SbD (spot-billed duck), Te (common teal), Wb (wild bird) and Wwf (wild waterfowl). Geographic locations are: FJ (Fujian), GD (Guangdong), GZ (Guizhou), HK (Hong Kong), JX (Jiangxi), RZ (Rizhao), SD (Shandong), SH (Shanghai), WZ (Wenzhou) and ZJ (Zhejiang). Viruses from different hosts are indicated by: humans, circles; chickens, blue names; domestic ducks or geese, green names.

However, our data show that, for this gene, more closely related H11N9 and H2N9 viruses are found in migratory wild birds in Hong Kong in 2010–2011 (Fig. 1b, asterisk). Phylogenetic linkage between the viruses of these birds and those of domestic ducks in China can be observed at least twice (Fig. 1b, asterisk) before the emergence of the N9 gene in the 2013 outbreak (Fig. 1b and Supplementary Fig. 3).

These data also allow us to reconstruct the genesis of the novel H7N7 lineage closely related to the H7N9 outbreak strain. The N7 of this lineage (Fig. 1c, lineage 'd') arose from H7N7 viruses present in the domestic waterfowl of China since at least 2010 (Fig. 1c and Supplementary Fig. 4). After entering domestic ducks, these early H7N7 viruses and H7N3 viruses (which had internal genes from a sub-lineage, ZJ-5, of the wild bird viral gene pool) co-circulated in these birds in eastern China during 2010–2011 (Fig. 1) and gave rise to H7N7 viruses with ZJ-5 internal genes (WZ-Dk-H7N7; Fig. 1a). H7N7 viruses were also introduced to chickens and by reassortment obtained internal genes from co-circulating H9N2 viruses, generating the H7N7 viruses found in Wenzhou (WZ-Ck-H7N7; Figs 1 and 2).

The H9N2 viruses that contributed internal genes to the H7N9 and H7N7 lineages were formed by reassortment of the major H9N2 lineage in China (SH-F/98) with a Eurasian wild bird virus, from which the former acquired a polymerase basic 2 (PB2) segment, creating a separate sub-lineage (ZJ-HJ/07; Fig. 2 and Supplementary Figs 5–10). The internal genes of the ZJ-HJ/07 sub-lineage form two distinct subgroups (Supplementary Figs 5–10, marked 'α' and 'β'), based on the PB2 segment phylogeny. Reassortment between α and β led to the internal gene cassettes of the H7N9 and H7N7 viruses having an NS gene from α and other internal genes from β (Fig. 2).

Even though the H7N9 and H7N7 chicken isolates obtained their internal gene cassettes from similar H9N2 viruses, all of their six internal genes formed distinct sublineages by subtype (Fig. 1d and Supplementary Figs 5–10). This indicates that the H7N9 and H7N7 viruses resulted from two independent interspecies transmissions to chickens, most likely from domestic ducks, and subsequent reassortment events. Some H7N9 and H7N7 viruses had one or two internal genes that fell outside these sublineages and instead clustered with other co-circulating H9N2 chicken viruses (Supplementary Figs 5–10). Thus, further reassortments with other H9N2 viruses continued to occur in chickens, increasing the genetic diversity of the H7N9 and H7N7 viruses.

Both the H7N9 and H7N7 lineages have three major sources of their eight gene segments (Fig. 2). Although the H7N7 virus suggests a path of two consecutive reassortment events (Fig. 2, yellow arrows), the evidence is not as clear for H7N9 viruses (Fig. 2, purple arrows). Two models are supported for the generation of the H7N9 outbreak lineage by molecular clock dating of divergence times: either a single reassortment event between mid-2011 and mid-2012, or a reassortment between H7 and N9 viruses in 2011 to generate an H7N9 ancestor, probably in ducks, followed by an interspecies transmission and a second reassortment with chicken H9N2 viruses from mid-2011 to late 2012 (Supplementary Fig. 11; substitution rates are shown in Supplementary Table 2). However, alternative reassortant pathways and hosts cannot be definitively excluded due to the limited numbers of H7 and N9 viruses identified during outbreak surveillance, and it is possible for either the H7 or N9 viruses to have first reassorted with H9N2 viruses.

The H7N9 and H7N7 viruses have similar, but independent evolutionary origins. Their haemagglutinin genes originated from H7 viruses that have been introduced to and established among the domestic ducks in China since 2010. At that time, these duck H7 viruses had internal genes from the ZJ-5 sub-lineage, as did H3N3, H1N2 and H11N3 duck viruses found in southern and eastern China during 2009–2013 (Fig. 1). An N9 virus that was a related precursor of the H7N9 viruses (Dk/JX/21714/11; H11N9), had polymerase genes from the ZJ-5 sub-lineage (Fig. 1b, asterisk; Fig. 1d), linking duck N9 viruses to the duck H7N3 precursors of the H7N9 lineage.

The haemagglutinin of the H7N9 outbreak lineage (including viruses isolated from Rizhao, but excluding Shanghai/1/2013 and Ck/ZJ/DTID/ZJU01/2013) had the amino acid substitution Gln235Leu/Ile (position 226 in H3 numbering) indicating that this substitution, which favours α2,6 sialic acid binding, seems to have originated in chickens. The equivalent substitution in chicken H9N2 viruses has been observed since the late 1990s and was linked to virus replication in the upper respiratory tract of birds[6]. The haemagglutinins of the H7N7 lineage from Wenzhou, and their closely related duck viruses (Fig. 1a), do not have this substitution. Of the other amino acid substitutions linked to α2,6 sialic acid binding seen in the H7N9 viruses (Gly195Val, Ala146Ser; 186, 138 in H3 numbering)[4], only Gly195Val is seen in one H7N7 virus (Supplementary Fig. 12). However, two recent studies have shown that, despite these mutations, the H7 haemagglutinin has limited binding to human receptors[8,9]. Both the neuraminidase proteins of the H7N7 and H7N9 chicken viruses have overlapping deletions (20 amino acids in N7, positions 53–72; five amino acids in N9, positions 69–73) in the stalk region, which are often observed in influenza virus lineages established in terrestrial poultry[10]. Duck H7 viruses (N7, N3 and N2) from Wenzhou did not have a stalk deletion. The Glu627Lys and Asp701Asn substitutions in PB2 normally seen in mammalian viruses[11] are only found in human H7N9 isolates (but not in all) and not in the chicken H7N9 and H7N7 viruses obtained here. Although it cannot be excluded, there is no evidence from these data to suggest that

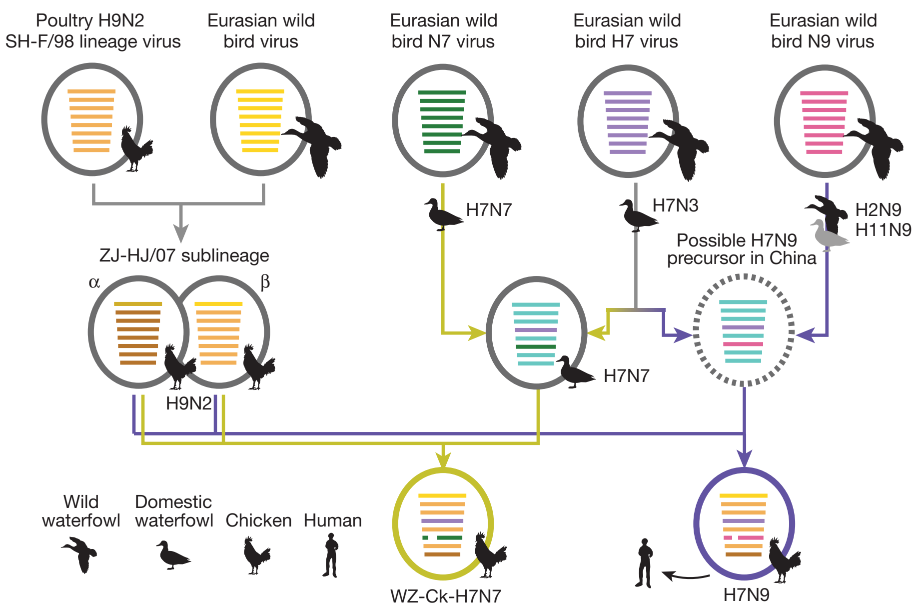


Figure 2 | Evolutionary pathways of the H7N9 and H7N7 viruses. Virus particles are represented by coloured ovals containing horizontal bars that represent the eight gene segments (from top to bottom: PB2, PB1, polymerase acidic, haemagglutinin, nucleoprotein, neuraminidase, matrix and non-structural). Segments in descendent viruses are coloured according to their corresponding source viruses (top) to illustrate gene ancestry through reassortment events. Source viruses for a reassortment are adjacent to arrow tails; arrowheads point to the resulting reassortants. Bars coloured cyan indicate gene segments of the ZJ-5 sub-lineage of wild bird viruses. A broken bar in segment 6 (neuraminidase) indicates a stalk region deletion. The virus indicated by a broken oval represents a hypothetical reassortant.

these viruses are mammalian adapted or that a mammalian intermediate host was involved in the human H7N9 infections.

The chicken H7N7 viruses carry only some of the molecular markers seen in the human H7N9 isolates (Supplementary Fig. 12), but they may still have the potential to infect humans or mammals. To assess infectivity in mammals, two groups ($n = 6$ each) of ferrets were inoculated with Chicken/Wenzhou/610/2013 (H7N7) at $10^{4.5}$ or $10^{6.5}$ plaque-forming units (p.f.u.). Both groups shed virus from 2 days post-inoculation (d.p.i.) (high dose) and 3 d.p.i. (low dose), and virus could be detected in rectal swabs at 4 d.p.i. in both groups (Supplementary Fig. 13). Infectious virus and positive nucleoprotein-stained cells could be detected in the nasal turbinate, and also in the trachea, lungs and hilar lymph nodes, with marginally higher levels from the high dose group at 3 and 5 d.p.i. (Supplementary Figs 13 and 14). This shows that the H7N7 lineage viruses can cause significant infection in mammals under experimental conditions, although virus shedding is lower than those of the 2009 pandemic H1N1 and 2013 human H7N9 viruses[12].

These findings provide a comprehensive picture of the creation and establishment of the H7N9 viruses that have infected humans. Domestic ducks seem to act as key intermediate hosts by acquiring and maintaining diverse influenza viruses from migratory birds, by facilitating the generation of different combinations of H7 and N9 or N7 subtype viruses, and by transmitting these viruses to chickens. After transmission, reassortment with enzootic H9N2 viruses formed the current H7N9 or H7N7 viruses seen in chickens. This probably led to outbreaks in chickens, resulting in the rapid spread of the novel reassortant H7N9 lineage through LPMs, which then became the source of human infections. The cessation of human infections after the closure of LPMs[13], after a precedent set during the Hong Kong H5N1 'bird flu' incident in 1997 (ref. 14), strongly supports this proposition.

The detection of H7N7 chicken viruses in Wenzhou that, like H7N9 viruses, have the potential to infect mammals suggests the current pandemic threat extends beyond the H7N9 virus. Even though human infections with the H7N9 virus seem to be under control, it is too early to know whether this virus has been eradicated from chickens over a larger geographic region. It is possible that H7N9 or H7N7 viruses are still present and may become enzootic in poultry. To control H7N9 and related viruses ultimately, it is necessary to reconsider the management of LPMs in urban areas. Long-term influenza surveillance remains essential for early warning of novel reassortant viruses and interspecies transmission events.

METHODS SUMMARY

Oropharyngeal, cloacal or faecal samples were taken from poultry at LPMs, farms and from wild birds at wetlands in regions flanking the recent H7N9 influenza outbreak. These samples were assessed for the presence of influenza viruses and, with archived H7 and N9 viruses from surveillance in southern China, subjected to full genome sequencing using Sanger or next-generation sequencing methods. After alignment of these sequences with available published sequence data (Supplementary Tables 3–5), maximum likelihood phylogenetic trees were constructed using the GTR+Γ model and topological robustness was assessed[15–17]. Molecular dating was undertaken in a Bayesian Markov chain Monte Carlo framework to calculate divergence times[18]. Ancestral sequences were reconstructed using maximum likelihood methods and codons undergoing positive selection were identified[19–21]. Ferrets were intranasally inoculated with Chicken/Wenzhou/610/2013 (H7N7) at high and low doses. Measurement of virus shedding, titration of virus and examination of pathology in major organs were undertaken as previously described[12].

Full Methods and any associated references are available in the online version of the paper.

Received 22 May; accepted 1 August 2013.
Published online 21 August 2013.

1. Gao, R. *et al.* Human infection with a novel avian-origin influenza A (H7N9) virus. *N. Engl. J. Med.* **368,** 1888–1897 (2013).
2. World Health Organization. Number of confirmed human cases of avian influenza A(H7N9) reported to World Health Organization; http://www.who.int/influenza/human_animal_interface/influenza_h7n9/08_ReportWebH7N9Number.pdf (2013).
3. Liu, D. *et al.* Origin and diversity of novel avian influenza A H7N9 viruses causing human infection: phylogenetic, structural, and coalescent analyses. *Lancet* **381,** 1926–1932 (2013).
4. Kageyama, T. *et al.* Genetic analysis of novel avian A(H7N9) influenza viruses isolated from patients in China, February to April 2013. *Euro Surveill.* **18,** 20453 (2013).
5. Hvistendahl, M., Normile, D. & Cohen, J. Influenza. Despite large research effort, H7N9 continues to baffle. *Science* **340,** 414–415 (2013).
6. Guo, Y. J. *et al.* Characterization of the pathogenicity of members of the newly established H9N2 influenza virus lineages in Asia. *Virology* **267,** 279–288 (2000).
7. Duan, L. *et al.* Influenza virus surveillance in migratory ducks and sentinel ducks at Poyang Lake, China. *Influenza Other Respi. Viruses* **5** (suppl. 1), 65–68 (2011).
8. Tharakaraman, K. *et al.* Glycan receptor binding of the influenza A virus H7N9 hemagglutinin. *Cell* **153,** 1486–1493 (2013).
9. Xiong, X. *et al.* Receptor binding by an H7N9 influenza virus from humans. *Nature* **499,** 496–499 (2013).
10. Cheung, C. L. *et al.* Establishment of influenza A virus (H6N1) in minor poultry species in southern China. *J. Virol.* **81,** 10402–10412 (2007).
11. Yamada, S. *et al.* Biological and structural characterization of a host-adapting amino acid in influenza virus. *PLoS Pathog.* **6,** e1001034 (2010).
12. Zhu, H. *et al.* Infectivity, transmission, and pathology of human-isolated H7N9 influenza virus in ferrets and pigs. *Science* **341,** 183–186 (2013).
13. Xu, J., Lu, S., Wang, H. & Chen, C. Reducing exposure to avian influenza H7N9. *Lancet* **381,** 1815–1816 (2013).
14. Shortridge, K. F. *et al.* Interspecies transmission of influenza viruses: H5N1 virus and a Hong Kong SAR perspective. *Vet. Microbiol.* **74,** 141–147 (2000).
15. Anisimova, M., Gil, M., Dufayard, J. F., Dessimoz, C. & Gascuel, O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst. Biol.* **60,** 685–699 (2011).
16. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59,** 307–321 (2010).
17. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–2690 (2006).
18. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29,** 1969–1973 (2012).
19. Murrell, B. *et al.* Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8,** e1002764 (2012).
20. Pond, S. L., Frost, S. D. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21,** 676–679 (2005).
21. Pupko, T., Pe'er, I., Shamir, R. & Graur, D. A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Mol. Biol. Evol.* **17,** 890–896 (2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank our colleagues from the Joint Influenza Research Centre (SUMC/HKU) and the State Key Laboratory of Emerging Infectious Diseases for their technical assistance. This study was supported by the National Institutes of Health (National Institute of Allergy and Infectious Diseases contract HSN266200700005C), Li Ka Shing Foundation, the Area of Excellence Scheme of the University Grants Committee of the Hong Kong SAR (grant AoE/M-12/06), Shenzhen Peacock Plan High-End Talents Program (KQTD201203), the University Development Fund (HKU) and the Innovation and Technology Commission of the Hong Kong Government. T.T.-Y.L. was supported in part by a Newton International Fellowship of the Royal Society. Metabiota's involvement was supported by the US Agency for International Development (USAID) Emerging Pandemic Threats Program, PREDICT project, under the terms of Cooperative Agreement Number GHN-A-OO-09-00010-00. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no. 278433-PREDEMICS, ERC Grant agreement no. 260864 and the Wellcome Trust (grant 092807) to A.R. and S.J.L.

Author Contributions Y.G., H.Z. and T.T.-Y.L. conceived the study; J.W., Y. Shen, B.Z., L.D., C.Y.-H.L., W.H., Z.O. and X.C. conducted surveillance; H.Z., J.W., C.-L.C., C.M., L.L., Y.C., L.Z., H.L., Y.L., A.F. and D.J.K. performed virus isolation, sequencing and animal experiments; T.T.-Y.L., A.R., O.G.P., H.Z., D.K.S., S.J.L., L.L.M.P., J.S.M.P., G.M.L., Y. Shu, R.G.W, R.J.W. and Y.G. contributed to the analysis; D.K.S. and T.T.-Y.L. wrote the manuscript; Y.G., H.Z., O.G.P. and A.R. edited the manuscript.

Author Information All sequences generated by this study have been deposited in GenBank under accession numbers KF258943–KF260956 and KF297287–KF297322 (Supplementary Table 3). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.Z. (zhuhch@hku.hk) or Y.G. (yguan@hku.hk).

METHODS

Influenza virus surveillance in Zhejiang, Shandong and Guangdong provinces. LPMs, poultry farms and a wetland were surveyed for influenza viruses. If accessible, paired oropharyngeal and cloacal samples were taken from birds, otherwise, samples from isolated individual faecal droppings were collected. Some drinking water samples were also collected. Two cities that flanked the initial outbreak around Shanghai and a more distant city were surveyed (Supplementary Fig. 1), as these locations might have viruses related to the outbreak and/or viruses similar to those that led to the genesis of the outbreak virus. Sampling was conducted immediately before the closure of markets and/or culling of the birds.

Sampling was conducted at LPMs, farms and a wetland in Wenzhou, Zhejiang, a city approximately 500 km south of Shanghai, from 7 to 10 April 2013. Poultry from LPMs and farms were sampled at Rizhao, Shandong, a city approximately 600 km north of Shanghai, on 17 and 18 April 2013. Sampling was also carried out at LPMs in Shenzhen, a major urban centre adjoining Hong Kong and approximately 1,200 km south of Shanghai, on 4, 6 and 7 April 2013. Details of the numbers and types of poultry sampled are given in Supplementary Table 1. After collection, each swab was placed in transport medium (M199) with antibiotics and kept in a cool box before and during shipping to the analysis laboratory. Swab materials were inoculated into 9–10-day-old embryonated chicken eggs and incubated for 48 h at 37 °C. Haemagglutinin-positive isolates were collected and further subtyped by haemagglutination and neuraminidase inhibition (HAI and NAI) assays using a panel of reference antisera as described previously[22]. Standard precautions were taken to avoid cross contamination of samples. Original samples positive for H7 were confirmed with rapid diagnostic quantitative PCR with reverse transcription (qRT–PCR) for the presence of H7 and N9 gene segments (see below). All H7 and H9 influenza isolates from the Wenzhou markets ($n = 78$, Supplementary Table 1), and all H7 isolates ($n = 8$) and 18 of the 51 H9 isolates from the Rizhao markets and farms were selected for full genome sequencing.

Selection of archived samples collected during 2000–2013. Archived H7 ($n = 293$, from 45 sampling occasions) and N9 ($n = 202$, from 27 sampling occasions) viruses isolated during 2000–2013 from our avian influenza surveillance conducted in Fujian, Guangdong, Guangxi, Guizhou, Jiangxi, Yunnan and Hong Kong (Supplementary Fig. 1) were isolated using embryonated eggs and identified by standard HAI and NAI assays as described above. A total of 42 N9 isolates (1 H1, 5 H2, 1 H3, 1 H4, 11 H7, 4 H10 and 19 H11) and a further 57 H7 (1 N1, 26 N3, 4 N6, 23 N7 and 3 N8) isolates were selected for full genome sequencing. For 63 sampling occasions, all H7 and N9 isolates were sequenced. For two occasions from Fujian, 2 out of 5 and 3 out of 6 isolates were sequenced. For six occasions from Jiangxi, 3 out of 5, 3 out of 7, 4 out of 10, 6 out of 11, 3 out of 33 and 10 out of 171 were sequenced. This was based on at least one virus from each HAI and NAI titre group (that is, those that had similar values from these assays). A further 98 H9N2 viruses isolated from Hong Kong retail poultry markets were sequenced.

Genomic sequencing. All selected isolates had all eight segments sequenced either using Sanger sequencing on an ABI 3730 genetic analyser (Applied Biosystems) or high-throughput next-generation sequencing on a Genome Sequencer Junior (Roche). Where a sample contained more than one copy of any segment (indicating a mixed infection of two viruses), only sequences of the haemagglutinin and neuraminidase sequences that could be explicitly defined or have been confirmed by single clonal Sanger sequencing were used in subsequent phylogenetic analyses.

Rapid diagnostic qRT–PCR. For rapid determination of the presence of any H7 and N9 viruses, primers specific for these subtypes were designed, based on the alignment of all H7 and N9 gene sequences available from GenBank on 1 April 2013. Viral RNA was extracted using the QIAamp Viral RNA Minikit (QIAGEN), and complementary DNAs were synthesized with primer Uni12 using the Transcriptor High Fidelity cDNA Synthesis Kit (Roche). Copy numbers of the H7 haemagglutinin or the N9 neuraminidase segment were determined on a LightCycler 480 Real-Time PCR System (Roche) using LightCycler FastStart DNA MasterPLUS SYBR Green I Kit (Roche). Primers were H7-765 (5′-GTTTCAATGGGGCHTTCATAGC-3′), H7-995R (5′-ACATTCTTCATCCCTGTWGC-3′), N9-160 (5′-CAAGCCAAACAATAATAAACAA-3′) and N9-293R (5′-CATGAATTTATAGTACAGAGYCCT-3′).

Sequence collection and alignment. All previously published influenza A virus sequences were collated from GenBank on 5 April 2013. Seven full genome sequences of H7N9 influenza viruses from the outbreak in China were downloaded from GISAID (http://gisaid.org) and two from GenBank (http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html). Acknowledgement of the sources of the GISAID sequences is given in Supplementary Table 4. These database sequences were combined with the sequences generated in this study (Supplementary Table 3) for further analysis. Sequences of each gene segment were aligned using MUSCLE v3.5 (ref. 23) with manual adjustment.

Phylogenetic analysis. Preliminary global phylogenetic analysis of all sequences was done using RAxML v7.6.8 (ref. 17). This identified the major Eurasian lineage that contains the H7N9 human isolates and the H7N7 and H7N9 isolates sequenced here. The major Eurasian lineage was reduced to approximately 1,000 sequences. Care was taken to ensure all sequences closely related to the H7N9 outbreak isolates and the novel H7N7 viruses were retained (Supplementary Table 5). Maximum likelihood phylogenies were reconstructed using the GTR+Γ nucleotide substitution model in PhyML v3.0 (ref. 16). Topological robustness was assessed by a Shimodaira–Hasegawa approximate likelihood ratio (SH-like) branch test[15]. The phylogenies for H7, N9, N7, PB2, PB1, polymerase acidic, nucleoprotein, matrix and non-structural segments are shown in Supplementary Figs 2–10. Smaller trees, with only essential taxa are presented in Fig. 1 with bootstrap support values from 1,000 pseudo-replicates to indicate the confidence of the topology.

Molecular dating. The evolutionary timescales of the H7N7 and H7N9 phylogenies were inferred by a molecular clock model that was calibrated using the virus sampling dates. Gene sequences of the H7N9 and H7N7 viruses were analysed with a selected set of closely related viruses. A relaxed clock model with uncorrelated lognormal distribution[24], tree topologies and other evolutionary parameters were jointly estimated using the Bayesian Markov chain Monte Carlo method implemented in BEAST v1.7.5 (ref. 18). The SRD06 nucleotide substitution model was used[25]. A Bayesian skyride model with time-aware smoothing was used[26] and a uniform prior and random walk operator was assigned to virus taxa whose sampling dates were known only to the nearest year. Multiple MCMC trajectories were computed and combined, giving from 2×10^8 to 10×10^8 total steps for each data set, with sampling every 1,000 steps. Convergence, at effective sample sizes >200, of relevant parameters was assessed in Tracer v1.5 (http://tree.bio.ed.ac.uk/software/tracer/). Maximum clade credibility trees with branches scaled by time were summarized, and the time of divergence and time of most recent common ancestor of the H7N9 and H7N7 viruses were obtained as illustrated in Supplementary Fig. 11. All data sets were screened to exclude mosaic sequences, as previously described[27], before molecular clock and further evolutionary analyses were undertaken.

Genome substitutions. The ancestral nucleotide sequence at each internal node of the maximum likelihood tree was reconstructed by the maximum likelihood joint method[21] implemented in HYPHY v2.0 (ref. 20). The sequences at both ends of a tree branch were compared, and the differences represent the substitutions that occurred along that branch. Substitutions in the positively selected codons identified by the mixed effects model of evolution[19] in Datamonkey[28] are highlighted in Supplementary Fig. 12.

Animal infection experiments. Four- to five-month-old female Angora ferrets (*Mustela putorius furo*) were obtained through a laboratory ferret breeding program at the Wuxi Sangosho and confirmed to be influenza virus free by virus isolation in Madin–Darby canine kidney cells (ATCC) from nasal washes and rectal swabs and sero-negative by HAI assay against contemporary swine and human influenza viruses (swine: H1 and H3; human: seasonal H1N1, H3N2 and influenza B), and avian H5N1 and H9N2 that are enzootic in China, and the chicken H7N7 virus to be tested in this study. A minimum number of animals necessary to obtain reproducible results were used according to the ethical guidelines. Ferrets were randomly allocated across the treatment groups, to which investigators were not blinded.

A group of six ferrets were intranasally inoculated with a dose of $10^{6.5}$ p.f.u. (or $10^{6.8}$ 50% tissue culture infectious dose ($TCID_{50}$)) of the Chicken/Wenzhou/610/2013 (H7N7) virus (high dose group) and a second group were inoculated with $10^{4.5}$ p.f.u. (or $10^{4.8}$ $TCID_{50}$) of viruses (low dose group). Nasal and rectal washes were taken daily from each individual and titrated using standard $TCID_{50}$ assays. At 3 and 5 d.p.i., three ferrets from each group were euthanized and tissues from each major organ (pulmonary lobes, nasal turbinate, upper and lower trachea and hilar lymph nodes) were collected for virus titration ($TCID_{50}$ assays), RNA extraction and qRT–PCR, and studies of tissue pathology (haematoxylin and eosin and viral nucleoprotein staining), following protocols described previously[12]. All animal experiment protocols were reviewed and approved by the Institutional Ethical Review Board (IERB) of Shantou University Medical College (ref no. SUMC2013-111) and The University of Hong Kong Committee on the Use of Live Animals in Teaching & Research. All experiments with H7N7 and H7N9 viruses were conducted in biosafety level 3 (BSL3) facilities, using enhanced BSL3 practices for the animal work and following practices in the approved institutional guidelines.

22. Huang, K. *et al.* Establishment and lineage replacement of H6 influenza viruses in domestic ducks in southern China. *J. Virol.* **86,** 6075–6083 (2012).
23. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
24. Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4,** e88 (2006).
25. Shapiro, B., Rambaut, A. & Drummond, A. J. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol. Biol. Evol.* **23,** 7–9 (2006).
26. Minin, V. N., Bloomquist, E. W. & Suchard, M. A. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* **25,** 1459–1471 (2008).

27. Lam, T. T. *et al.* Systematic phylogenetic analysis of influenza A virus reveals many novel mosaic genome segments. *Infect. Genet. Evol.* **18,** 367–378 (2013).

28. Delport, W., Poon, A. F., Frost, S. D. & Kosakovsky Pond, S. L. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26,** 2455–2457 (2010).

doi:10.1038/nature12526

Type 2 innate lymphoid cells control eosinophil homeostasis

Jesse C. Nussbaum[1], Steven J. Van Dyken[1], Jakob von Moltke[1], Laurence E. Cheng[2], Alexander Mohapatra[3], Ari B. Molofsky[4], Emily E. Thornton[5], Matthew F. Krummel[5], Ajay Chawla[1,6,7], Hong-Erh Liang[1] & Richard M. Locksley[1,3,8]

Eosinophils are specialized myeloid cells associated with allergy and helminth infections. Blood eosinophils demonstrate circadian cycling, as described over 80 years ago[1], and are abundant in the healthy gastrointestinal tract. Although a cytokine, interleukin (IL)-5, and chemokines such as eotaxins mediate eosinophil development and survival[2], and tissue recruitment[3], respectively, the processes underlying the basal regulation of these signals remain unknown. Here we show that serum IL-5 levels are maintained by long-lived type 2 innate lymphoid cells (ILC2) resident in peripheral tissues. ILC2 cells secrete IL-5 constitutively and are induced to co-express IL-13 during type 2 inflammation, resulting in localized eotaxin production and eosinophil accumulation. In the small intestine where eosinophils and eotaxin are constitutive[4], ILC2 cells co-express IL-5 and IL-13; this co-expression is enhanced after caloric intake. The circadian synchronizer vasoactive intestinal peptide also stimulates ILC2 cells through the VPAC2 receptor to release IL-5, linking eosinophil levels with metabolic cycling. Tissue ILC2 cells regulate basal eosinophilopoiesis and tissue eosinophil accumulation through constitutive and stimulated cytokine expression, and this dissociated regulation can be tuned by nutrient intake and central circadian rhythms.

Eosinophils require survival signals delivered through the common β-receptor chain (βc) shared by IL-3, IL-5 and granulocyte–macrophage colony-stimulating factor (GM-CSF)[1]. IL-5 is particularly important, as supported by studies in IL-5-deficient[5] and IL-5 receptor α chain (IL-5Rα)-deficient[6] mice, and in humans using anti-IL-5 and anti-IL-5Rα antibodies that target eosinophils in disease. Without IL-5 signalling, residual eosinophils have been attributed to IL-3 and GM-CSF, as well as eosinophil chemokines, such as eotaxins, that sequester these cells into tissues[1,3].

To identify cells that support eosinophils, we generated IL-5 reporter mice, designated Red5 (recombinase-expressing detector for IL-5; R5)[7]. Cells from these mice contain a tandem dimer red fluorescent protein (tdTomato) linked by an internal ribosomal entry site (IRES) to a Cre element replacing the translation initiation site of the endogenous *Il5* gene, facilitating function marking, fate mapping and deletion based on IL-5 expression (Fig. 1a). We validated that the construct disrupts the endogenous *Il5* gene and that R5 fluorescence correlates with IL-5 production using $CD4^{+}$ T cells examined after T-helper type 1 (T_H1) or T-helper type 2 (T_H2) polarization (Extended Data Fig. 1a–c).

Consistent with previous observations[8], $CD45^{+}R5^{+}CD4^{-}$ cells were present in non-lymphoid tissues, including the brain, heart, lung, kidney, skin, intestine and uterus, whereas few $R5^{+}$ cells were in lymphoid organs, including spleen, bone marrow, lymph nodes or thymus, or in the liver (Fig. 1b, c). In all tissues, the vast majority of $R5^{+}$ cells were small cells (forward-/side-scatter low) that lacked lineage markers for T, B, natural killer (NK) and myeloid cells, and expressed markers for innate lymphoid type 2 cells (ILC2)[9], including CD90.2 (Thy1), CD127 (IL-7Rα), KLRG1 and ICOS; T1/ST2 and CD25 expression varied among tissues

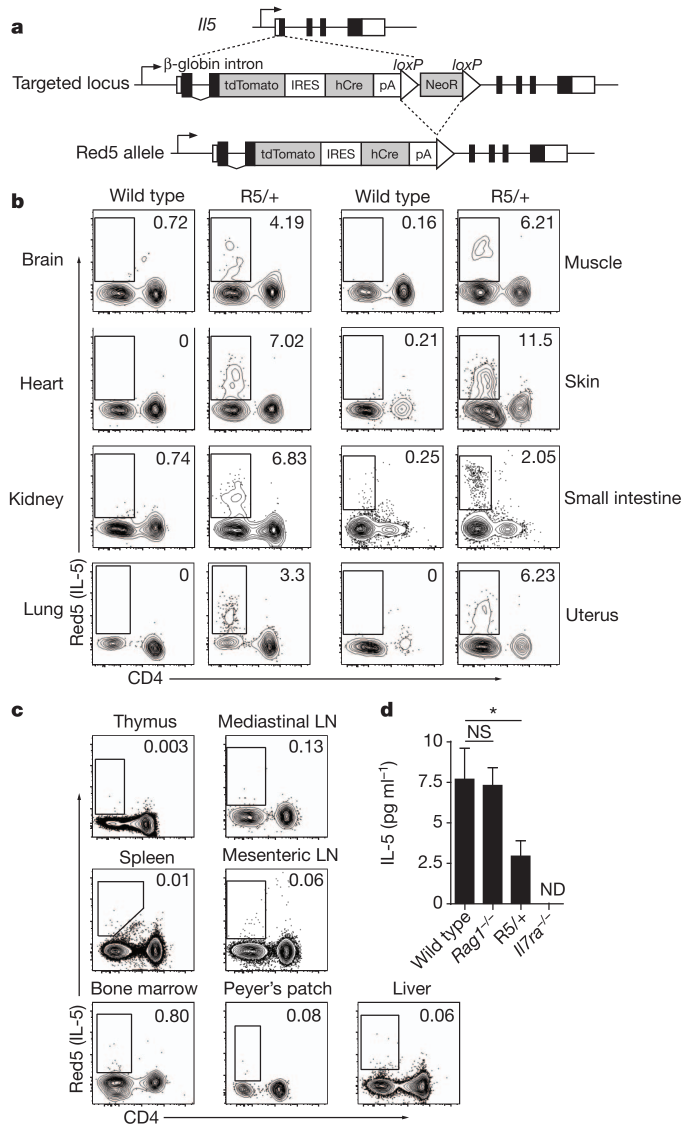


Figure 1 | Innate cells produce IL-5 in tissues at rest. **a**, Schematic of targeting construct. **b**, **c**, Flow cytometry of tissues, previously gated on $CD45^{+}CD90.2^{+}$ cells in wild-type and R5/+ (**b**) or $CD90.2^{+}$ cells in R5/R5 (**c**) naive mice. **d**, Serum IL-5. Data are representative of two independent experiments with two mice per group (**b**–**d**) or pooled from three independent experiments for 7 (wild type), 4 (Red5), or 8 (others) mice per group (**c**). Represented as mean ± s.e.m. LN, lymph nodes; ND, none detected; NS, not significant; $*P < 0.05$ by Student's *t*-test.

[1]Department of Medicine, University of California San Francisco, San Francisco, California 94143-0795, USA. [2]Department of Pediatrics, University of California San Francisco, San Francisco, California 94143-0795, USA. [3]Departments of Microbiology & Immunology, University of California San Francisco, San Francisco, California 94143-0795, USA. [4]Department of Laboratory Medicine, University of California San Francisco, San Francisco, California 94143-0795, USA. [5]Department of Pathology, University of California San Francisco, San Francisco, California 94143-0795, USA. [6]Department of Physiology, University of California San Francisco, San Francisco, California 94143-0795, USA. [7]Cardiovascular Research Institute, University of California San Francisco, San Francisco, California 94143-0795, USA. [8]Howard Hughes Medical Institute, University of California San Francisco, San Francisco, California 94143-0795, USA.

but was present on all lung R5$^+$ ILC2 cells (Extended Data Fig. 2a). Most (75–80%) CD90.2$^+$T1/ST2$^+$ lung ILC2 cells were R5$^+$, and these cells had the highest reporter expression compared to the few ILC2 cells from bone marrow and lymph nodes (Extended Data Fig. 2b). Sorted R5/+ ILC2 cells spontaneously secreted IL-5 in culture, confirming that the reporter marks IL-5 production (Extended Data Fig. 2c). Consistent with previous reports[9], numbers of ILC2 and R5$^+$ ILC2 cells were similar in recombinase activating gene (RAG)-deficient mice but were nearly absent in CD127-deficient mice (Extended Data Fig. 3a–c). Correspondingly, serum IL-5 was comparable in wild-type and RAG-deficient mice, reduced to about half-normal levels in heterozygous R5/+ mice, and was not detected in CD127-deficient mice (Fig. 1d).

Few ILC2 cells were present in the lungs of newborn mice, but within the first week CD90.2$^+$T1/ST2$^+$ cells increased, and the percentage that were R5$^+$ reached adult levels (Fig. 2a and Extended Data Fig. 3d–e). We administered 5-bromodeoxyuridine (BrdU) in drinking water of adult mice for 2 weeks to label dividing cells and found that less than 10% of lung ILC2 cells were labelled (Extended Data Fig. 3f). Pulse–chase labelling indicated that the decay of labelled ILC2 cells was substantially slower than CD4$^+$ T cells (Fig. 2b). As assessed using multiphoton microscopy, lung R5$^+$ ILC2 cells were embedded in collagen-rich regions near the confluence of medium-sized blood vessels and airways but were absent from alveoli (Fig. 2c and data not shown).

Lung eosinophilia is a hallmark of allergic lung disease and helminth migration, but eosinophils are rare in the lung at baseline[1] despite constitutive local IL-5. In previous studies, ILC2 cells stimulated with cytokines or helminth infection upregulated IL-13 (refs 10, 11), which is genetically linked to IL-5 in mice and humans and induces epithelial eotaxins (including CCL11) and endothelial adhesins necessary for eosinophil trafficking[1,3]. We crossed R5 mice to Smart13 (also called *Il13*$^{tm2.1Lky}$) reporter mice, in which non-signalling human CD4 marks cells producing IL-13 (ref. 11). In contrast to resting ILC2 cells, lung ILC2 cells expressed the IL-13 reporter after infection with the helminth *Nippostrongylus brasiliensis*. All IL-13$^+$ ILC2 cells in the lung were R5$^+$, whereas CD4$^+$ T cells expressed IL-5, IL-13, or both cytokines, consistent with previous observations (Fig. 3a)[11].

We crossed R5/R5 mice to mice carrying a ROSA26-flox stop-YFP allele to fate-map cells that expressed the IL-5-linked Cre recombinase, and infected the mice with *N. brasiliensis* to elicit a type 2 immune response. After infection, YFP was present only in ILC2 cells and CD4$^+$ T cells, and all YFP$^+$ cells were also R5$^+$ (Extended Data Fig. 4a, b). We also crossed R5/R5 mice to mice carrying a ROSA26-flox stop-diphtheria toxin A allele to delete IL-5-producing cells. The R5 allele was designed such that expression of the tdTomato reporter precedes Cre-mediated *loxP* recombination. Therefore, in R5/R5 deleter mice, a population of R5^{lo} cells may be detectable before they express the ROSA26-diphtheria toxin. At baseline, R5/R5 and R5/R5 deleter mice had comparable numbers of total cells and CD4$^+$ T cells in the bone marrow, spleen, lung and small intestine lamina propria (Extended Data Fig. 4c), but R5$^+$ ILC2 cells were deleted in the lung (Fig. 3b) and small intestine (Extended Data Fig. 4d).

To study the activity of lung ILC2 cells in the absence of T_H2 cells, we crossed R5/R5 and R5/R5 deleter mice onto a RAG-deficient background and administered IL-2 and IL-33 (refs 10, 12, 13). As expected, cytokine-activated ILC2 cells in R5/R5 RAG-deficient mice showed increased surface KLRG1 expression and R5 mean fluorescence intensity (Extended Data Fig. 5). Cytokine administration increased the ILC2 population and induced eotaxin-1 (CCL11) in lungs of R5/R5 RAG-deficient mice but not in RAG-deficient R5/R5 deleter mice (Fig. 3c). ILC2 cell deficiency was bypassed by administering IL-13, which partially restored eotaxin levels.

Whereas eosinophils are rare in the lung at baseline, they are abundant in other tissues, such as the small intestine lamina propria, where they depend on CCL11 (ref. 4), and are absent in mice that lack ILC2 cells (refs 13, 14). Our finding that ILC2 cells in the lung can control

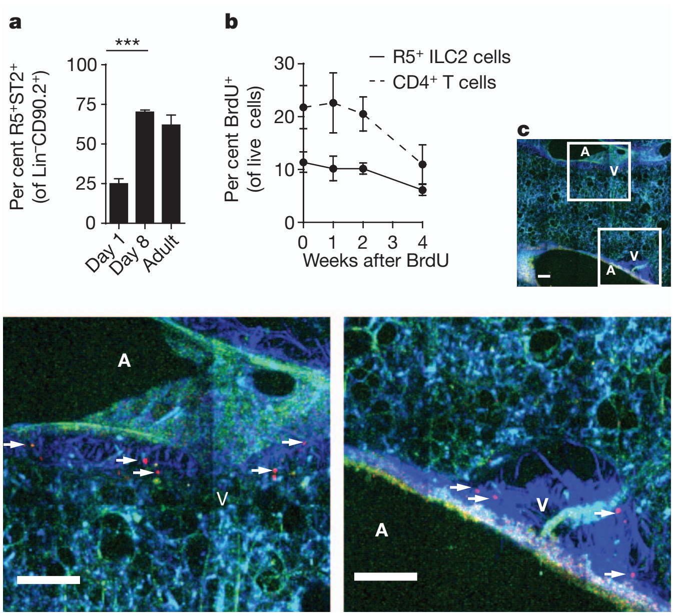


Figure 2 | ILC2 cells expand after birth and persist in collagen-rich structures. **a**, Per cent of lung Lin$^-$CD90.2$^+$ cells R5$^+$T1/ST2$^+$ on day 1, day 8, or week 8. **b**, Per cent BrdU$^+$ of R5$^+$ ILC2 cells and total CD4$^+$ cells in lung after 4 weeks BrdU. **c**, Representative multiphoton images of tdTomato fluorescence (red) in naive R5/R5 actin–CFP mice; CFP and autofluorescence in blue and green, respectively. A, airway; V, vasculature. Collagen second harmonic appears blue. Scale bars, 100 µm. Data are pooled from three independent experiments for 5 (day 1), 6 (day 8), or 4 (adult) mice per group (**a**); or pooled from two independent experiments for 5 (week 0), 6 (week 1), or 3 (others) mice per group (**b**). Represented as mean ± s.e.m. Images represent eight regions taken from two mice. Lin, lineage markers (B220, CD5, CD11b, CD11c, Ly6G, FcεRI and NK1.1); ***$P < 0.01$ by Student's *t*-test.

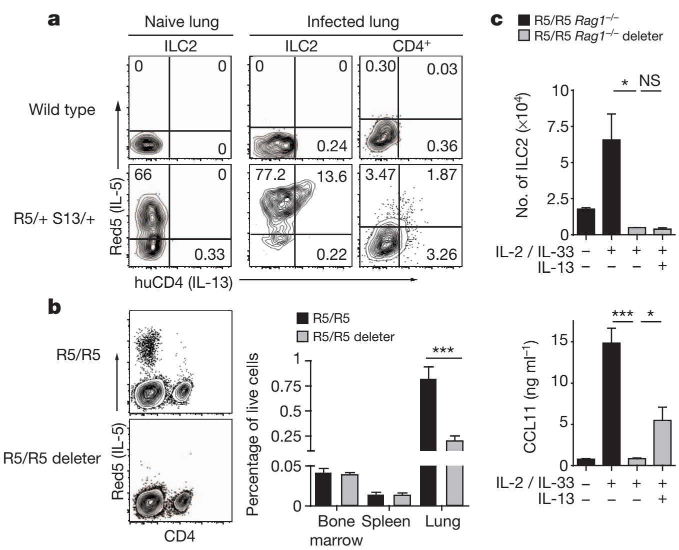


Figure 3 | IL-5 and IL-13 co-expression in lung ILC2. **a**, Lung IL-5 and IL-13 reporter expression before and after infection. **b**, Flow cytometry of CD90.2$^+$ lung cells and per cent with ILC2 surface markers (CD90.2 and either KRLG1, T1/ST2 or CD25) at rest. **c**, ILC2 (left lung) and CCL11 concentration (right lung) after IL-2, IL-33 and IL-13 treatment. Data are representative of three independent experiments with 4 (naive R5$^+$Smart13$^+$ (S13$^+$)), 5 (infected R5$^+$Smart13$^+$), or 2 (others) mice per group (**a**), pooled from three independent experiments for 6 (R5/R5 bone marrow and spleen) or 9 (all others) mice per group (**b**), or pooled from two independent experiments for 8 (R5/R5 + IL-2/IL-33), 5 (R5/R5 deleter + IL-2/IL-33/IL-13), or 3 (others) mice per group (**c**). Represented as mean ± s.e.m. huCD4, human CD4; NS, not significant; *$P < 0.05$; ***$P < 0.001$, by Student's *t*-test.

eosinophil recruitment through dissociated expression of IL-5 and IL-13 led us to explore the role of ILC2-derived cytokines in the basal regulation of peripheral eosinophils. We measured serum IL-5 levels at 10:00 and at 22:00 and found that the levels correlated with the circadian variation in blood eosinophils[15]. Although influenced by the adrenal–cortical axis[15], blood eosinophils can be dominantly synchronized by meal timing[16]. Mice fasted for 16 h showed suppressed serum IL-5 and blood eosinophils at 10:00 (Fig. 4a, b). To minimize the effects of altered light/dark cycle or stress induced by fasting, we restricted two groups of mice to feeding at night only or day only, allowed both groups to acclimatize for 9 days, and analysed them at 8:00 (Extended Data Fig. 6a). Unlike lung ILC2 cells, lamina propria ILC2 cells express IL-13 constitutively, and this was increased in the morning if mice had just been in a fed as opposed to fasted cycle (Fig. 4c and Extended Data Fig. 6b); the numbers of ILC2 cells remained constant (Extended Data Fig. 6c). IL-13 reporter expression by intestinal ILC2 at 8:00 was diminished after a 16-h overnight fast (Extended Data Fig. 6d) and restored by administering an evening food (but not water) gavage at 22:00 (Fig. 4d).

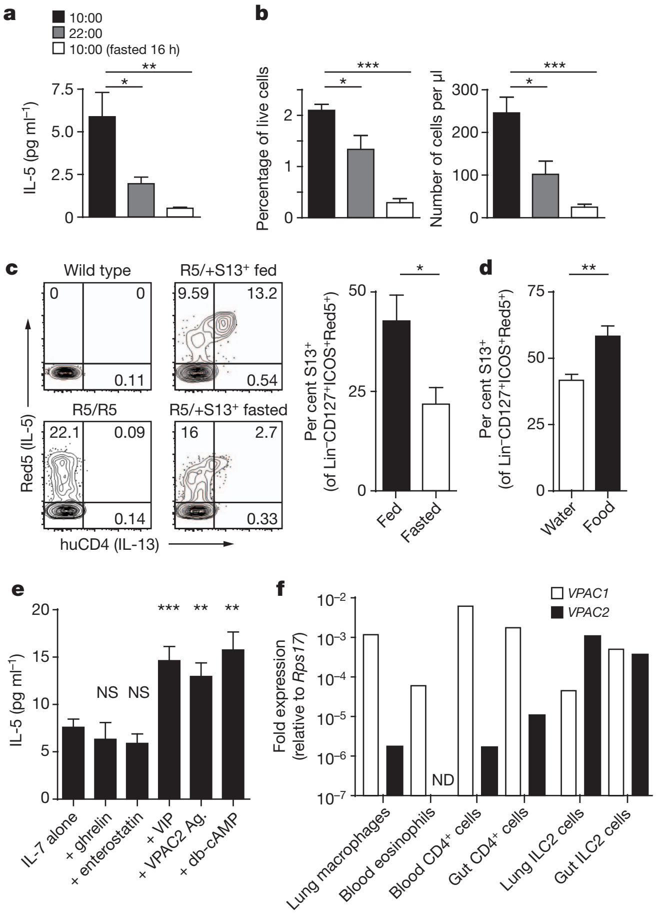


Figure 4 | ILC2 cells respond to circadian and metabolic cues. **a**, **b**, Serum IL-5 and blood eosinophils at 10:00, 22:00 or at 10:00 after fasting. **c**, **d**, Flow cytometry of small intestine ILC2 cells (Lin$^-$CD127$^+$ICOS$^+$) and percentage of R5^{hi} ILC2 cells expressing Smart13 at 8:00 in mice on night-time (black) or daytime (white) feeding (**c**) or in fasted mice given food (black) or water (white) by oral gavage (**d**). **e**, Supernatant IL-5 from intestinal Lin$^-$CD45$^+$KLRG1$^+$ ILC2 cells cultured in IL-7 alone or with indicated reagents. **f**, Expression of *VPAC1* and *VPAC2* in sorted cells, relative to *Rps17*. Data pooled from independent experiments for 19 (AM), 6 (PM), or 5 (fasted) mice per group (**a**); 7 (AM), 4 (PM), or 8 (fasted) mice per group (**b**); 8 mice per group (**c**); 6 mice per group (**d**); or pooled averages of duplicate cultures from 6 (IL-7 alone, plus VIP, plus VPAC2 agonist) or 3 (all others) cell sorts from independent mice (**e**), or representative of two experiments of independent cell sorts (**f**). Represented as mean ± s.e.m. Lin, lineage markers (B220, CD11b, CD11c, Ly6G, FcεRI and NK1.1); Rps17, 40S ribosomal protein S17; NS, not significant; $*P < 0.05$; $**P < 0.01$; $***P < 0.001$, by Student's t-test. db-cAMP, dibutyryl-cAMP.

The response of small intestine ILC2 cells to caloric input raised the possibility that these cells could respond to hormonal cues induced by feeding. Indeed, purified intestinal ILC2 cells, most of which were R5$^+$ (Extended Data Fig. 7a), released detectable IL-5 when incubated with IL-7 alone, but increased IL-5 production with addition of vasoactive intestinal peptide (VIP) but not ghrelin or enterostatin (Fig. 4e and Extended Data Fig. 7b). The effect on lung ILC2 cells was similar (Extended Data Fig. 7c). VIP is a member of the secretin family of neuropeptides, which are expressed throughout the nervous system. They are highly expressed in intestinal neurons, coordinating pancreatic secretion with smooth muscle relaxation in response to feeding[17], and in neurons of the suprachiasmatic nucleus, relaying environmental cues necessary to synchronize central circadian oscillators[18]. VIP receptor type 2 (VPAC2)-deficient and VIP-deficient mice show similar defects in circadian behaviour[19,20], and rhythms mediated by VPAC2 are entrained by feeding[21].

We proposed that VIP might signal through VPAC2 on ILC2 cells. VIP and its receptors are also expressed by immune cells, and signals through VPAC2 have been implicated in T_H2 cell expansion, survival and cytokine production[22]. Mice deficient in VPAC2 have delayed infiltrating eosinophils with allergic challenge[23]. We detected both *VPAC1* (also called *Vipr1*) and *VPAC2* (*Vipr2*) messenger RNA in intestinal and lung ILC2 cells, whereas *VPAC2* expression was undetectable in eosinophils and low in macrophages and CD4$^+$ T cells (Fig. 4f). Comparable levels of IL-5 were induced in culture with VIP or with a VPAC2-specific agonist (Fig. 4e and Extended Data Fig. 7c)[24]. VPAC2 is a G-protein-coupled receptor that can associate with $G\alpha_s$ to activate adenylate cyclase[25]. Consistent with this, we also induced comparable IL-5 levels bypassing the receptor with dibutyryl-cAMP (Fig. 4e).

First noted in humans over 80 years ago[26], circadian variation of blood eosinophils has been linked to neuroendocrine[15] and metabolic[16] cycling. As shown here, long-lived ILC2 cells in peripheral tissues are the predominant source of circulating IL-5, and their close association with vasculature positions these cells for eosinophil recruitment. After stimulation by epithelial and/or T_H2 cytokines, lung ILC2 cells increase IL-5 and co-express IL-13, leading to local eosinophil accumulation, a process that mimics the post-prandial response of intestinal ILC2 cells to caloric intake. Furthermore, ILC2 cells express functional VPAC2 receptors, providing a potential mechanism linking these dispersed tissue-resident cells with central circadian and metabolic rhythms. Intestinal eosinophils are normal in germ-free animals[4], and IL-13$^+$ ILC2 cells are found in human fetal gut[27], indicating that these biological pathways are independent of intestinal microbiota. Although further study is needed, our findings indicate that eosinophils are linked to basal circadian oscillations through ILC2 cell activation and raise the possibility that helminthic parasites may have co-opted these fundamental pathways of host metabolic homeostasis.

METHODS SUMMARY

IL-5 reporter mice were generated as described[28], all mice on a C57BL/6 background. Tissues were mechanically dissociated, and lungs, heart, kidney, uterus, brain, intestine, skin and muscle were digested in Liberase (Roche) with or without DNase I (Roche). Some tissues were separated by 90/40 Percoll gradient (GE Healthcare). Cells were stained and analysed on an LSR II (BD) or sorted on a MoFlo XDP (Beckman Coulter). ELISA (R&D Systems) was used to detect IL-5 and CCL11, or IL-5 was detected using Enhanced Sensitivity Cytometric Bead Array (BD). BrdU was detected using a BrdU flow kit (BD Biosciences). Multiphoton imaging was performed as described[29]. Infection with 500 *N. brasiliensis* L3 larvae was as described[11]. Blood for IL-5 and eosinophils was collected as a terminal procedure. For gavage, fasted mice received water or 1:1 high-fat chow (Research Diets)[7] plus 20% dextrose at 22:00. *VPAC1* and *VPAC2* transcripts were quantified by RT–PCR[30]. All data were analysed using Prism (GraphPad Software). Paired two-tailed Student's t-tests were used for BrdU experiments and for *in vitro* IL-5; Kruskal–Wallis was used to compare multiple groups (Extended Data Fig. 3c); otherwise comparisons were made with unpaired two-tailed Student's t-tests applying Welch's correction as indicated.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 November 2012; accepted 2 August 2013.

Published online 15 September 2013.

1. Rothenberg, M. E. & Hogan, S. P. The eosinophil. *Annu. Rev. Immunol.* **24,** 147–174 (2006).
2. Takatsu, K. & Nakajima, H. IL-5 and eosinophilia. *Curr. Opin. Immunol.* **20,** 288–294 (2008).
3. Pope, S. *et al.* IL-13 induces eosinophil recruitment into the lung by an IL-5- and eotaxin-dependent mechanism. *J. Allergy Clin. Immunol.* **108,** 594–601 (2001).
4. Mishra, A., Hogan, S., Lee, J., Foster, P. & Rothenberg, M. Fundamental signals that regulate eosinophil homing to the gastrointestinal tract. *J. Clin. Invest.* **103,** 1719–1727 (1999).
5. Kopf, M. *et al.* IL-5-deficient mice have a developmental defect in $CD5^+$ B-1 cells and lack eosinophilia but have normal antibody and cytotoxic T cell responses. *Immunity* **4,** 15–24 (1996).
6. Yoshida, T. *et al.* Defective B-1 cell development and impaired immunity against *Angiostrongylus cantonensis* in IL-5Rα-deficient mice. *Immunity* **4,** 483–494 (1996).
7. Molofsky, A. B. *et al.* Innate lymphoid type 2 cells sustain visceral adipose tissue eosinophils and alternatively activated macrophages. *J. Exp. Med.* **210,** 535–549 (2013).
8. Ikutani, M. *et al.* Identification of innate IL-5-producing cells and their role in lung eosinophil regulation and antitumor immunity. *J. Immunol.* **188,** 703–713 (2012).
9. Spits, H. & Cupedo, T. Innate lymphoid cells: emerging insights in development, lineage relationships, and function. *Annu. Rev. Immunol.* **30,** 647–675 (2012).
10. Price, A. *et al.* Systemically dispersed innate IL-13-expressing cells in type 2 immunity. *Proc. Natl Acad. Sci. USA* **107,** 11489–11494 (2010).
11. Liang, H. E. *et al.* Divergent expression patterns of IL-4 and IL-13 define unique functions in allergic immunity. *Nature Immunol.* **13,** 58–66 (2012).
12. Neill, D. R. *et al.* Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity. *Nature* **464,** 1367–1370 (2010).
13. Moro, K. *et al.* Innate production of TH2 cytokines by adipose tissue-associated c-Kit^+Sca-1^+ lymphoid cells. *Nature* **463,** 540–544 (2010).
14. Carlens, J. *et al.* Common γ-chain-dependent signals confer selective survival of eosinophils in the murine small intestine. *J. Immunol.* **183,** 5600–5607 (2009).
15. Halberg, F., Visscher, M. & Bittner, J. Eosinophil rhythm in mice: range of occurrence; effects of illumination, feeding, and adrenalectomy. *Am. J. Physiol.* **174,** 109–122 (1953).
16. Pauly, J. *et al.* Meal timing dominates the lighting regimen as a synchronizer of the eosinophil rhythm in mice. *Acta Anat.* **93,** 60–68 (1975).
17. Lelievre, V. *et al.* Gastrointestinal dysfunction in mice with a targeted mutation in the gene encoding vasoactive intestinal polypeptide: a model for the study of intestinal ileus and Hirschsprung's disease. *Peptides* **28,** 1688–1699 (2007).
18. Maywood, E. S. *et al.* Analysis of core circadian feedback loop in suprachiasmatic nucleus of mCry1-luc transgenic reporter mouse. *Proc. Natl Acad. Sci. USA* **110,** 9547–9552 (2013).
19. Harmar, A. J. *et al.* The $VPAC_2$ receptor is essential for circadian function in the mouse suprachiasmatic nuclei. *Cell* **109,** 497–508 (2002).
20. Colwell, C. S. *et al.* Disrupted circadian rhythms in VIP- and PHI-deficient mice. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **285,** R939–R949 (2003).
21. Sheward, W. J. *et al.* Entrainment to feeding but not to light: circadian phenotype of VPAC2 receptor-null mice. *J. Neurosci.* **27,** 4351–4358 (2007).
22. Voice, J. *et al.* c-Maf and JunB mediation of Th2 differentiation induced by the type 2 G protein-coupled receptor (VPAC2) for vasoactive intestinal peptide. *J. Immunol.* **172,** 7289–7296 (2004).
23. Samarasinghe, A. E., Hoselton, S. A. & Schuh, J. M. The absence of the VPAC2 receptor does not protect mice from *Aspergillus* induced allergic asthma. *Peptides* **31,** 1068–1075 (2010).
24. Tsutsumi, M. *et al.* A potent and highly selective VPAC2 agonist enhances glucose-induced insulin release and glucose disposal: a potential therapy for type 2 diabetes. *Diabetes* **51,** 1453–1460 (2002).
25. Dickson, L. & Finlayson, K. VPAC and PAC receptors: From ligands to function. *Pharmacol. Ther.* **121,** 294–316 (2009).
26. Domarus, A. v. Die bedeutung der kammerzahlung der eosinophilen fur die klinik. *Dtsch. Arch. Klin. Med.* **171,** 333–358 (1931).
27. Mjösberg, J. M. *et al.* Human IL-25- and IL-33-responsive type 2 innate lymphoid cells are defined by expression of CRTH2 and CD161. *Nature Immunol.* **12,** 1055–1062 (2011).
28. Sullivan, B. M. *et al.* Genetic analysis of basophil function *in vivo*. *Nature Immunol.* **12,** 527–535 (2011).
29. Thornton, E. E. *et al.* Spatiotemporally separated antigen uptake by alveolar dendritic cells and airway presentation to T cells in the lung. *J. Exp. Med.* **209,** 1183–1199 (2012).
30. Vomhof-DeKrey, E. E. *et al.* Radical reversal of vasoactive intestinal peptide (VIP) receptors during early lymphopoiesis. *Peptides* **32,** 2058–2066 (2011).

Acknowledgements We thank the NIH Tetramer Core Facility for reagents, B. Sullivan, N. Flores, M. Consengco and Z. Wang for technical expertise, and M. Anderson, C. Lowell and M. McCune for comments on the manuscript. Supported by NIH (AI026918, AI030663, AI078869, HL107202), the Diabetes Endocrinology Research Center grant (DK063720), the Howard Hughes Medical Institute and the Sandler Asthma Basic Research Center at the University of California San Francisco. J.C.N. is supported by NIH training grants (AI007641 and AI007334).

Author Contributions J.C.N. performed experiments, interpreted data and wrote the manuscript; L.E.C. provided experimental and imaging assistance; S.J.V.D., A.M., A.B.M. and J.v.M. provided experimental assistance; E.E.T. performed imaging assays; M.F.K. provided reagents and expertise; H.-E.L. generated mouse cytokine reporter strains; A.C. discussed experiments and provided oversight for metabolic studies; R.M.L. directed the studies and wrote the paper with J.C.N.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.M.L. (locksley@medicine.ucsf.edu).

LETTER

doi:10.1038/nature12488

Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication

Atsuya Nishiyama[1], Luna Yamaguchi[1], Jafar Sharif[2], Yoshikazu Johmura[1], Takeshi Kawamura[3], Keiko Nakanishi[4], Shintaro Shimamura[5], Kyohei Arita[6], Tatsuhiko Kodama[3], Fuyuki Ishikawa[7], Haruhiko Koseki[2] & Makoto Nakanishi[1]

Faithful propagation of DNA methylation patterns during DNA replication is critical for maintaining cellular phenotypes of individual differentiated cells[1–5]. Although it is well established that Uhrf1 (ubiquitin-like with PHD and ring finger domains 1; also known as Np95 and ICBP90) specifically binds to hemi-methylated DNA through its SRA (SET and RING finger associated) domain and has an essential role in maintenance of DNA methylation by recruiting Dnmt1 to hemi-methylated DNA sites[6–10], the mechanism by which Uhrf1 coordinates the maintenance of DNA methylation and DNA replication is largely unknown. Here we show that Uhrf1-dependent histone H3 ubiquitylation has a prerequisite role in the maintenance DNA methylation. Using *Xenopus* egg extracts, we successfully reproduce maintenance DNA methylation *in vitro*. Dnmt1 depletion results in a marked accumulation of Uhrf1-dependent ubiquitylation of histone H3 at lysine 23. Dnmt1 preferentially associates with ubiquitylated H3 *in vitro* though a region previously identified as a replication foci targeting sequence[11]. The RING finger mutant of Uhrf1 fails to recruit Dnmt1 to DNA replication sites and maintain DNA methylation in mammalian cultured cells. Our findings represent the first evidence, to our knowledge, of the mechanistic link between DNA methylation and DNA replication through histone H3 ubiquitylation.

Several lines of evidence showing the ability of Uhrf1 to bind modified histone H3 (refs 12–16) and the inability of Dnmt1 to catalyse its reaction on Uhrf1-bound hemi-methylated DNA[12,17] suggest that there may be an intermediate that connects Uhrf1 with Dnmt1 at DNA replication sites. Although direct interaction between Uhrf1 and Dnmt1 has been proposed as a means of recruiting Dnmt1 to hemi-methylated DNA sites[6,18,19], we failed to detect a stable interaction between Uhrf1 and Dnmt1 (Supplementary Fig. 2a–d). Therefore, to elucidate the mechanism underlying cooperation of maintenance DNA methylation and DNA replication by Uhrf1, we first reconstituted maintenance DNA methylation using *Xenopus* interphase egg extracts. After DNA replication, genomic DNA purified from sperm chromatin showed a marked incorporation of a radiolabelled methyl group when *S*-[methyl-^{3}H]-adenosyl-L-methionine was used as a donor, indicating that the sperm DNA was methylated (Fig. 1a). The addition of known replication inhibitors, such as aphidicolin, p27 or geminin, almost completely abolished DNA methylation as well as DNA replication (Fig. 1a, Supplementary Fig. 3a, and data not shown), indicating that DNA-replication-dependent DNA methylation can be reproduced in this *in vitro* cell-free system.

Although neither *Xenopus* (x)Uhrf1 nor xDnmt1 bound to mitotic chromatins, both began binding to chromatin with the kinetics similar to that of xPcna after mitotic extracts were stimulated to initiate DNA replication, suggesting a strong correlation between the chromatin loadings of xUhrf1 and xDnmt1 and DNA replication (Supplementary Fig. 3b, c). Treatment with aphidicolin and p27 abolished the chromatin loading of xDnmt1 and xUhrf1, although a slight loading and DNA-replication-independent binding of xUhrf1 to chromatin was still detectable.

Anti-xUhrf1 serum efficiently depleted xUhrf1 protein in the egg extract (Supplementary Fig. 4a). Depletion of xUhrf1 severely impaired xDnmt1 recruitment to the chromatin and DNA-replication-dependent DNA methylation, although xUhrf1 depletion did not affect DNA replication efficiency and timing when compared with mock depletion (Fig. 1b and Supplementary Fig. 4b, c). Add-back of recombinant human (rh)Uhrf1 partly rescued the recruitment of xDnmt1 to the chromatin and DNA-replication-dependent DNA methylation in xUhrf1-depleted extracts (Supplementary Fig. 4b and 5a). It should be noted that the depletion of xUhrf1 or xDnmt1 from the egg extract as well as mock depletion delayed DNA replication timing (see Supplementary Figs 3a and 4c, d), presumably owing to the dilution of interphase egg extracts by the immunodepletion procedure, which is usually observed in this *in vitro* cell-free system[20]. However, it did not affect the efficiency of DNA replication and chromatin loading of xPcna, indicating that neither xUhrf1 nor xDnmt1 was essential for DNA replication under our experimental condition (Fig. 1b and Supplementary Fig. 4c, d).

xDnmt1 depletion in extracts did not inhibit but rather enhanced xUhrf1 binding to chromatin, presumably owing to the increase in the hemi-methylated DNA region (Fig. 1c). Intriguingly, we detected slower-migrating forms of *Xenopus* histone H3 (xH3) in the xDnmt1-depleted chromatin fraction. These slower-migrating forms were hardly detectable in mock-depleted extracts, presumably owing to the rapid conversion of hemi-methylated DNA to fully methylated DNA by xDnmt1 (see also Fig. 1f). Immunodepletion of xUhrf1 from the extracts completely abolished the appearance of these slower-migrating forms in the xDnmt1-depleted extract (Fig. 1c), although DNA replication was not affected by double depletion of xDnmt1 and xUhrf1 (Supplementary Fig. 4d). Add-back of rhUhrf1 partly rescued the appearance of these slower-migrating forms of xH3 (Supplementary Fig. 5b). A band ladder with a higher molecular weight was detected in anti-H3 immunoprecipitates by anti-ubiquitin antibodies (Fig. 1d). In addition, we added an extreme excess of recombinant His-tagged ubiquitin to the egg extracts. In this case, we expected an upshift of the modified bands of xH3. As shown in Supplementary Fig. 6, an upshift of these bands was indeed apparent. Taken together, the slower-migrating forms that appeared after xDnmt1 depletion were ubiquitylated xH3 (xUbH3). Given that xUhrf1 is required for ubiquitylation of xH3 and that it possesses a RING finger domain (see Fig. 1c and Supplementary Fig. 5b), Uhrf1 is most likely to be an E3 ubiquitin ligase for H3. In agreement with this, Uhrf1 was reported to be able to ubiquitylate H3 *in vitro*[21–23].

Mass spectrometry analyses for xUbH3 in xDnmt1-depleted extracts were performed (Supplementary Fig. 7a). Tandem mass spectrometry (MS/MS) database searches of xUbH3 found one potentially modified

[1]Department of Cell Biology, Graduate School of Medical Sciences, Nagoya City University, 1 Kawasumi, Mizuho-cho, Mizuho-ku, Nagoya 467-8601, Japan. [2]RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro, Tsurumi-ku, Yokohama 230-0045, Japan. [3]Laboratory for System Biology and Medicine, RCAST, University of Tokyo, Komaba 4-6-1, Meguro-ku, Tokyo 153-8904, Japan. [4]Department of Perinatology, Institute for Developmental Research, Aichi Human Service Center, 713-8 Kamiya-cho, Kasugai, Aichi 480-0392, Japan. [5]Department of Molecular Medicine and Biochemistry, Akita University Graduate School of Medicine, 1-1-1 Hondo, Akita 010-8543, Japan. [6]Graduate School of Medical Life Sciences, Yokohama City University, 1-7-29, Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan. [7]Department of Gene Mechanisms, Graduate School of Biostudies, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan.

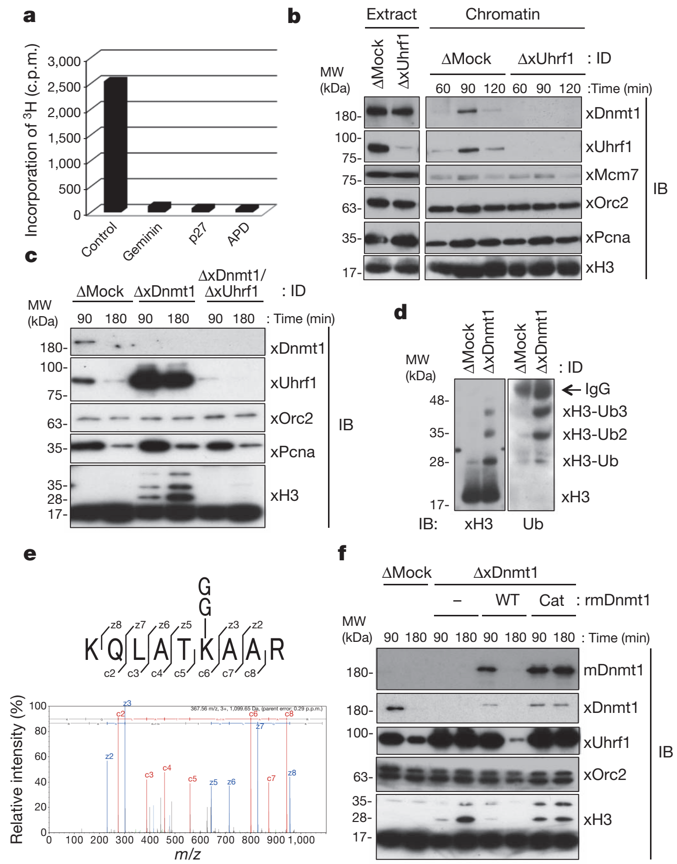


Figure 1 | xUhrf1- and DNA-replication-dependent DNA methylation and ubiquitylation of H3 at lysine 23 in *Xenopus* egg extracts. a, Sperm chromatin was incubated with interphase egg extracts containing radiolabelled *S*-[methyl-^{3}H]-adenosyl-L-methionine. His$_6$–geminin, glutathione *S*-transferase–p27 or aphidicolin (APD, 150 µM) was added to the extracts 10 min before the addition of sperm DNA. After 1 h, sperm DNA was purified from the egg extracts and the incorporation of radioactivity was measured using a scintillation counter. Data shown are representative of three independent experiments. **b**, Interphase egg extracts immunodepleted (ID) with the indicated antiserum were incubated with sperm nuclei. Extracts and sperm chromatin isolated at the indicated times were analysed by immunoblotting (IB) using the indicated antibodies. **c**, Chromatin fractions from interphase extracts immunodepleted with the indicated antibodies were isolated at the indicated times after sperm addition and analysed by immunoblotting. **d**, Chromatin fractions from the indicated immunodepleted extracts were solubilized by MNase and immunoprecipitated with anti-H3 antibodies after treatment with 1% SDS. The resultant immunoprecipitates were subjected to immunoblotting using the indicated antibodies. **e**, The electron transfer dissociation MS/MS spectrum of the xH3 peptide shows ubiquitylation at lysine 23. Detailed information is provided in Methods. **f**, Buffer alone (−), recombinant His$_8$-tagged wild-type (WT) rmDnmt1, or its catalytic mutant (cat) proteins (10 ng µl^{-1}) were added to xDnmt1-depleted extracts. Chromatin fractions isolated at the indicated times were subjected to immunoblotting using the indicated antibodies.

site at lysine 23 in a trypsin-digested sample[24] (Fig. 1e, see also Supplementary Fig. 7b, c). Lysine 23 in mammalian H3 has been previously reported following a systemic and quantitative proteome analysis of ubiquitylated proteins[25], suggesting that it is present *in vivo* and is physiologically relevant.

xUhrf1-dependent xH3 ubiquitylation in xDnmt1-depleted chromatin fractions was inhibited when DNA replication was blocked by the addition of p27 or aphidicolin (Supplementary Fig. 8), although trace but detectable amounts of xUhrf1 were detected on the chromatin in the presence of p27 or aphidicolin, indicating that ubiquitylation of xH3 by xUhrf1 requires on-going DNA replication.

Either recombinant wild-type or catalytic mutant (C1229S) mouse Dnmt1 (rmDnmt1) proteins[26,27], which were highly homologous to xDnmt1 (91% similarity), were reintroduced in xDnmt1-depleted extracts. Wild-type rmDnmt1 suppressed ubiquitylation of xH3, whereas C1229S failed to do so. Interestingly, wild-type, but not C1229S mutant, rmDnmt1 disappeared on chromatin at 180 min, concomitant with the loss of xUbH3, suggesting that the molecular coupling of deubiquitylation of xUbH3 and rmDnmt1 release from chromatin. These results indicate that recruitment of rmDnmt1 to chromatin per se is insufficient for prevention of xH3 ubiquitylation (Fig. 1f).

Anti-xH3 immunoprecipitates from xDnmt1-depleted extracts specifically bound to xDnmt1, but not to xUhrf1 (Fig. 2a and Supplementary Fig. 9a). The reciprocal experiment using immune-purified xDnmt1, xUhrf1 and xUsp7, a known Dnmt1-binding protein, also revealed that xDnmt1, but not xUhrf1 or xUsp7, bound to ubiquitylated xH3 (Supplementary Fig. 9b and 10a). We then examined whether the interaction of xDnmt1 with xUbH3 was direct. Anti-xH3 immunoprecipitates from xDnmt1-depleted chromatin bound more rmDnmt1 than those from mock-depleted chromatin (Fig. 2b). Specific interaction between rmDnmt1 and xUbH3 was also observed when rmDnmt1 protein was precipitated using anti-Flag antibodies (Fig. 2c). We then performed Far-western blotting analysis using rmDnmt1 protein as a probe. rmDnmt1 specifically bound to xUbH3, but not to other ubiquitylated proteins (see Fig. 2d, middle panel) or polyubiquitin chain per se, showing a preference for diubiquitylated xH3 (Fig. 2d), further confirming the direct and specific binding of Dnmt1 to UbH3. Although xDnmt1 immunoprecipitates bound to nucleosomal xH3 and xUbH3 (Supplementary Fig. 10a), rmDnmt1 preferentially bound to nucleosomal xUbH3 (Fig. 2c) and to monomeric di- or tri-ubiquitylated xH3 on the membrane (Fig. 2d). Although the structural basis underlying Dnmt1 binding to UbH3 is largely unknown, these somewhat distinctive observations might be explained by the possibility that xDnmt1 complexes contain proteins capable of binding to non-ubiquitylated H3 and by the distinct forms of H3, such as nucleosomes, core histones or histone monomer.

The targeting sequence in mouse (m)Dnmt1 that is required for association with replication foci has been previously identified[11]. Wild-type rmDnmt1 effectively bound to xUbH3, whereas a mutant rmDnmt1 lacking association with replication foci (ΔRep; rmDnmt1(Δ325–425)) failed to do so (Fig. 2e). Far-western analysis further confirmed the specific binding of wild-type, but not ΔRep, rmDnmt1 to xUbH3 (Supplementary Fig. 10b).

A similar human (h)Uhrf1-dependent ubiquitylation of hH3 was observed in HeLa cells. When hDnmt1 was depleted, ladders of slow-migrating bands of hH3 were readily detected (Fig. 3a). This ladder was also detected by anti-ubiquitin antibodies (Fig. 3b). However, this hUbH3 was not detected in hDnmt1 and hUhrf1 double-depleted cells (Fig. 3a). As with *Xenopus* egg extract (Supplementary Fig. 8), a ubiquitylation ladder of hH3 was specifically detected in S phase, but not in G1/S phase, chromatin (Supplementary Fig. 11). Notably, the ubiquitylation ladder of hH3 was detected even in the presence of hDnmt1 during normal S phase when hH3 immunoprecipitated from acid-extracted histones was used for immunoblotting (Fig. 3c). When *hDnmt1* was knocked down, the ectopically expressed hUbH3 was detectable by anti-Flag antibody in acid-extracted chromatin proteins from cells expressing wild-type, but not the K23R mutant, hH3 (Fig. 3d), confirming that the ubiquitylation in mammalian cells occurred at the same site as that observed in *Xenopus* extracts. We then examined whether lysine 23 ubiquitylation is prerequisite for Dnmt1 binding to nucleosomes. Ectopically expressed wild-type or K23R hH3 proteins in cells depleted of hDnmt1 were immunoprecipitated and incubated with purified rmDnmt1. The ubiquitylation ladder was only detected in wild-type hH3 precipitates that specifically interacted with rmDnmt1 (Fig. 3e). Taken together with the results using *Xenopus* extracts, Uhrf1-dependent ubiquitylation of H3 at

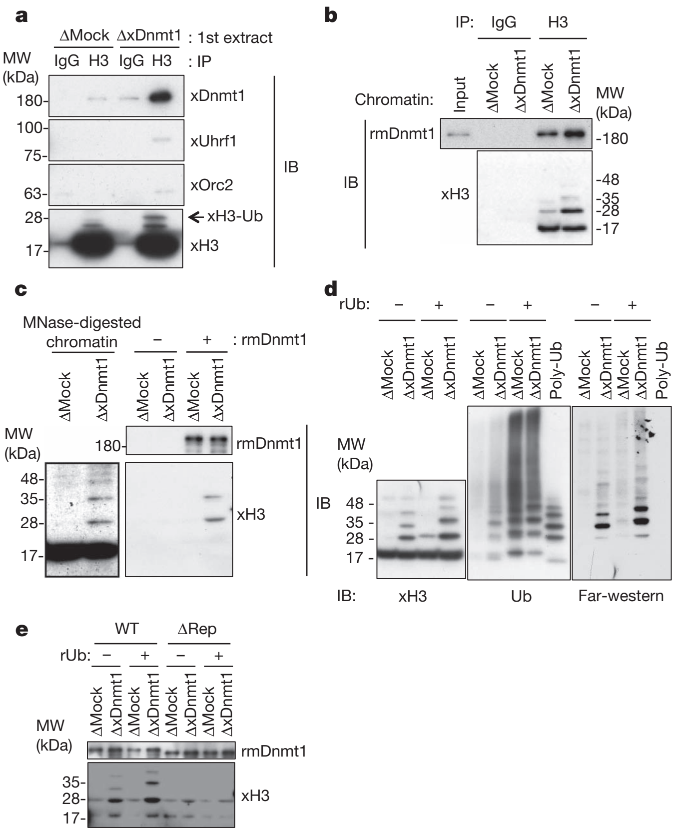


Figure 2 | Dnmt1 preferentially binds to UbH3. **a**, Chromatin fractions isolated from mock or xDnmt1-depleted extracts (first extracts) were solubilized by MNase and immunoprecipitated with anti-H3 antibodies after treatment with 1% SDS to increase the immunoprecipitation efficiency. The resultant immunoprecipitates were further incubated with untreated interphase extracts and the final immunoprecipitates were subjected to immunoblotting using the indicated antibodies. See also Supplementary Fig. 9a. **b**, Anti-H3 immunoprecipitates from either mock or xDnmt1-depleted chromatin solubilized by MNase were incubated with purified Flag-tagged rmDnmt1. The resultant immunoprecipitates were analysed by immunoblotting using the indicated antibodies. **c**, Flag-tagged rmDnmt1 coupled to beads (+) or control beads (−) was incubated with MNase-digested chromatin isolated from mock or xDnmt1-depleted extracts (left) and the resultant complex was analysed by immunoblotting as in **b** (right). **d**, Far-western blotting analysis of xUbH3 using rmDnmt1 as a probe. Mock or xDnmt1-depleted extracts were incubated with sperm chromatin in the presence or absence of an excess amount of recombinant His_6-ubiquitin (rUb). The chromatin fractions and poly ubiquitin chain (poly-Ub) were electroblotted onto a membrane and probed with anti-H3 antibodies (left), anti-ubiquitin antibodies (middle) or Flag-tagged-purified rmDnmt1 (right). Localization of the probe was carried out by immunoblotting using anti-Flag antibody. **e**, MNase-digested chromatins were incubated with either wild-type or a mutant Flag-tagged rmDnmt1 lacking the first 100 residues of the replication foci targeting sequence (ΔRep; mDnmt1(Δ325–425)). The mixtures were then immunoprecipitated with anti-Flag antibodies and the resultant immunoprecipitates were subjected to immunoblotting using anti-Flag antibodies (top) or anti-H3 antibodies (bottom).

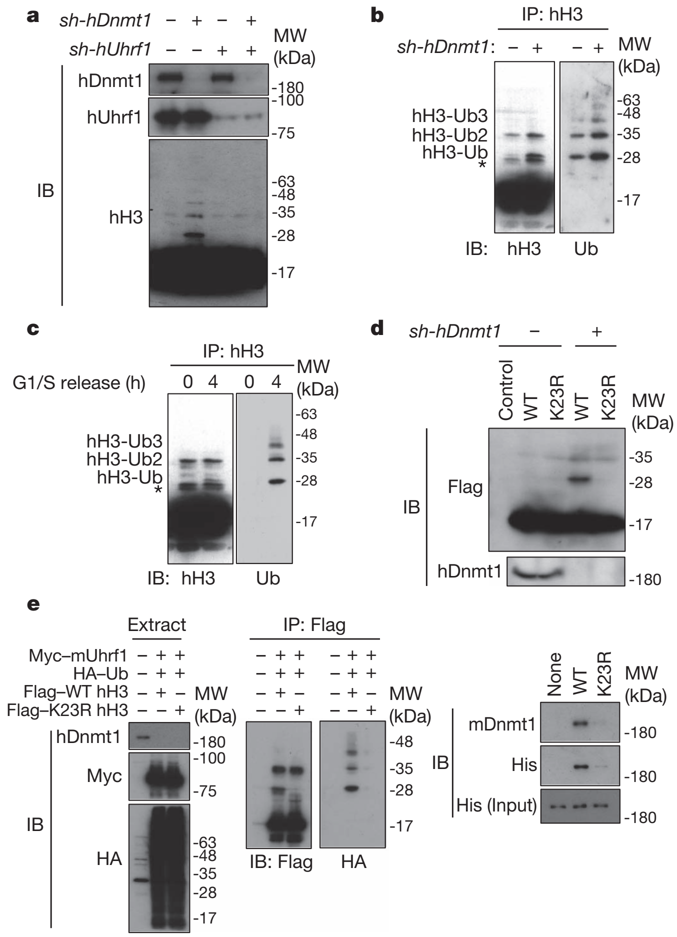


Figure 3 | hUhrf1- and S-phase-dependent ubiquitylation of hH3 at lysine 23 in HeLa cells. **a**, Extracts from asynchronous HeLa cells knocked down of control (−), *hDnmt1*, *hUhrf1* or *hDnmt1* and *hUhrf1* by their specific shRNAs were subjected to immunoblotting using the indicated antibodies. **b**, Acid-extracted histones from asynchronous HeLa cells knocked down for control (−) or *hDnmt1* were immunoprecipitated with anti-H3 antibodies. The resultant immunoprecipitates were subjected to immunoblotting using anti-H3 (left) and anti-Ub (right) antibodies. Asterisk indicates nonspecific bands. **c**, HeLa cells were synchronized at G1/S and released into S phase. Acid-extracted histones were subjected to immunoprecipitation using anti-H3 antibody. Immunoprecipitates were analysed by immunoblotting using anti-H3 (left) and anti-Ub (right) antibodies. Asterisk indicates nonspecific bands. **d**, *hDnmt1* was knocked down in HeLa cells ectopically expressing either wild-type or K23R mutant Flag-tagged hH3. Cells were collected at 72 h after depletion and histone proteins were isolated by an acid extraction. Cell extracts (bottom) and extracted histones (top) were subjected to immunoblotting using anti-Dnmt1 antibodies and anti-Flag antibodies, respectively. **e**, Ubiquitylation of hH3 at lysine 23 is required for nucleosomal binding of Dnmt1. HeLa cells were transfected with the indicated plasmids and the infected with lentiviruses expressing *sh-hDnmt1*. Cells were collected 72 h after infection and the extracts were immunoprecipitated with anti-Flag antibodies according to the experimental protocols from Abcam as described in Methods. After extensive wash, the immunoprecipitates were incubated with His_8-rmDnmt1 for 1 h at 4 °C. The final precipitates were washed four times and subjected to immunoblotting using the indicated antibodies (right) as well as the extracts (left) and Flag–immunoprecipitates (middle).

lysine 23 seems to be an essential and a general histone mark for recruitment of Dnmt1 in eukaryotes.

Finally, we determined the physiological significance of the E3 ubiquitin ligase activity of Uhrf1. We generated HeLa cells allowing a conditional replacement of endogenous hUhrf1 with either wild-type, SRA (D474G/R489A) or RING finger (C713A/C715A/C716A) mutants of mUhrf1. Wild-type mUhrf1 effectively restored the ability to ubiquitylate hH3 in cells knocked down for endogenous *hUhrf1*, whereas SRA and RING mutants failed to do so (Fig. 4a). Immunohistochemical analysis using $Uhrf1^{(\Delta/\Delta)}$ mouse embryonic fibroblasts ectopically expressing wild-type mUhrf1, SRA and RING mutants revealed that both mDnmt1 and mUhrf1 formed nuclear foci during S phase (Fig. 4b). These foci co-localized well with those of mPcna (Supplementary Fig. 12). Again, wild-type mUhrf1 restored the ability of mDnmt1 to form nuclear foci and co-localize with mPcna foci in cells in which endogenous mUhrf1 was deleted. By contrast, although the RING mutant itself formed nuclear foci (Fig. 4b and Supplementary Fig. 12), it failed to restore the ability of mDnmt1 to do so. As expected, the SRA

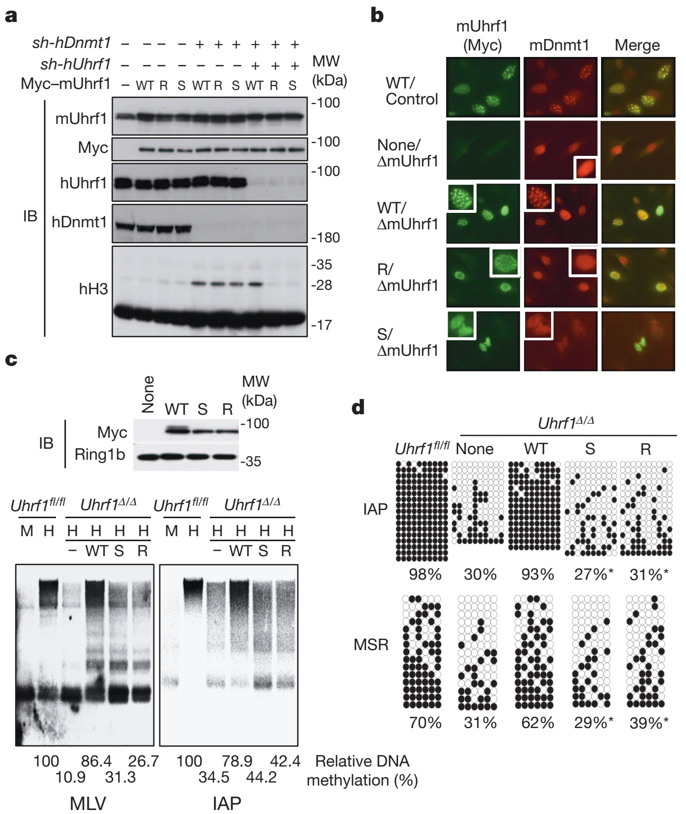


Figure 4 | The RING finger domain of Uhrf1 is crucial for recruitment of Dnmt1 to DNA replication sites and for maintenance of DNA methylation in mammalian cells. a, Endogenous *hDnmt1* and/or *hUhrf1* were knocked down from HeLa cells expressing various Myc-tagged mUhrf1 (RING finger domain mutant, R; SRA domain mutant, S) by their specific shRNAs. Cells were then collected 72 h after knockdown and the cell extracts were analysed by immunoblotting using the indicated antibodies. Note that antibodies against hUhrf1 failed to recognize mUhrf1 protein, whereas those against mUhrf1 recognized both proteins. **b**, Endogenous mUhrf1 was deleted from *Uhrf1*$^{fl/fl}$ mouse embryonic fibroblasts expressing various Myc-tagged mUhrf1 mutants by Cre recombinase expression. Cells were immunostained 72 h after deletion using anti-Myc (Myc–mUhrf1) and anti-Dnmt1 (mDnmt1) antibodies. Inserts show enlarged views of one nucleus. **c**, Endogenous mUhrf1 genes are conditionally deleted by the addition of 4-hydroxytamoxifen from *Uhrf1*$^{fl/fl}$ embryonic stem cells expressing ER-Cre. Global DNA methylation levels were examined in *Uhrf1* conditional knockout (*Uhrf1*$^{Δ/Δ}$) embryonic stem cells overexpressing the wild-type Uhrf1, the SRA domain mutant (S), the RING domain mutant (R) or control (−) in the presence (*Uhrf1*$^{Δ/Δ}$) or absence (*Uhrf1*$^{fl/fl}$) of 4-hydroxytamoxifen. The levels of protein expression were analysed by immunoblotting using anti-Myc antibodies and anti-Ring 1b antibodies as a control (top). Genomic DNA (1 μg) was treated with methylation-sensitive HpaII (H) or methylation-insensitive MspI enzyme (M) and analysed by Southern blotting using the indicated probes. Relative DNA methylation was determined by counting radioactivity of the undigested DNA (tops of the gels) and represented as a percentage of DNA undigested by HpaII in *Uhrf1*$^{fl/fl}$ mouse embryonic stem cells. **d**, Bisulphite DNA sequencing analysis was performed to detect the DNA methylation level in the Gag region of endogenous IAP retrotransposons and MSR (minor satellite repeats) in embryonic stem cells treated as in **c**. Closed circles indicate methylated CpGs and open circles represent unmethylated CpGs. The percentage of DNA methylation (methylation CpGs/total CpGs) is given at the bottom of each panel (*$P < 1 \times 10^{-6}$; compared to wild type).

mutant failed to form nuclear foci or to restore the ability of mDnmt1 to form them. These results suggest that the RING finger domain of mUhrf1 and consequent ubiquitylation of mH3 at lysine 23 are required for the recruitment of mDnmt1 to DNA replication sites. Similar results were also observed when mUhrf1 mutants in the conserved amino acids required for the putative interaction with E2 enzymes were used (Supplementary Fig. 13a–c)[28].

The level of DNA methylation in mouse embryonic stem cells was examined by DNA-methylation-sensitive Southern blotting using intracisternal A-type particle (IAP) and murine leukaemia virus (MLV) regions of the endogenous retrotransposon as probes. Wild-type mUhrf1 or SRA or RING mutants were introduced into *mUhrf1*$^{fl/fl}$ embryonic stem cells with their comparable expressions (Fig. 4c, top panels). The untreated embryonic stem cells showed a high level of DNA methylation in both IAP and MLV regions of the endogenous retrotransposon. When endogenous mUhrf1 was deleted, their methylation levels were significantly reduced, showing decreased amounts of undigested DNA by HpaII (Fig. 4c). Wild-type mUhrf1 effectively restored the level of global DNA methylation, whereas the RING mutant or the SRA mutant failed to do so. A bisulphite DNA sequencing analysis of the IAP Gag region of the endogenous retrotransposon and minor satellite repeats in mouse embryonic stem cells also demonstrated that wild-type mUhrf1, but not the RING mutant or the SRA mutant, restored the level of DNA methylation (Fig. 4d). Taken together, the results indicate that the RING finger domain of mUhrf1 and consequent ubiquitylation of mH3 at lysine 23 appear to be required for maintaining the level of DNA methylation in mouse embryonic stem cells.

In conclusion, our present results strongly suggest that Uhrf1-dependent ubiquitylation of H3 acts as a platform for the recruitment of Dnmt1 to DNA replication sites (Supplementary Fig. 1). The rapid turnover of H3 ubiquitylation during normal S phase has suggested the presence of a deubiquitylation mechanism coupled with Dnmt1-dependent DNA methylation rather than its proteasomal degradation. In fact, the Dnmt1 complex showed deubiquitylation activity towards ubiquitylated H3 (Supplementary Fig. 14a, b). Thus, it will be of great interest to elucidate the molecular basis underlying the coupling of this deubiquitylation- and Dnmt1-dependent DNA methylation.

METHODS SUMMARY

***Xenopus* egg extracts.** Cytostatic factor (CSF)-arrested and interphase egg extracts were prepared as described[29,30], except that metaphase II-arrested eggs were washed in extraction buffer (EB; 5 mM $MgCl_2$, 100 mM KCl, 20 mM HEPES-KOH, pH 7.5) instead of in 1× Marc's modified Ringers (MMR). Incubation of extracts was carried out at 22 °C.

Immunodepletion. For immunodepletion of *Xenopus* egg extracts, 250 μl of antiserum against xUhrf1 and xDnmt1 were conjugated to 75 μl of protein A agarose (Roche). Twenty-five microlitres of antibody beads were mixed with 100 μl interphase egg extracts and incubated at 4 °C for 1 h. This was repeated three times. For add-back experiments, mDnmt1 wild-type or the C1229S mutant was added to xDnmt1-depleted extracts at 10 ng μl^{-1}.

Monitoring DNA methylation and replication in *Xenopus* egg extracts. DNA methylation was monitored by the incorporation of *S*-[methyl-^{3}H]-adenosyl-L-methionine. Genomic DNA was purified using a Wizard Genomic DNA purification kit (Promega) according to the manufacturer's instructions. The DNA replication efficiency was determined as described previously[20].

Wild-type and mutant Uhrf1-expressing mouse embryonic stem cell cloning. Six micrograms of linearized pCAG-IRES-Puro plasmid containing either a Myc-tagged wild type, an SRA domain mutant (D474G, R489A) or a RING domain mutant (C713A,C715A,C716A) Uhrf1 complementary DNA were transfected into *Uhrf1*$^{fl/fl}$ embryonic stem cells using FuGENE HD transfection reagent (Roche Applied Science, no. 6366244). Stable clones were selected by puromycin (1 mg ml^{-1}) treatment.

Global DNA methylation analysis and bisulphite DNA sequencing. Genomic DNA was purified from embryonic stem cells using an AllPrep DNA/RNA mini kit according to the supplier's protocol. Genomic DNA was treated with HpaII or MspI and analysed by methylation-sensitive Southern blotting using MLV or IAP as probes. For bisulphite DNA sequencing analysis, genomic DNAs were treated with sodium bisulphite using an EpiTect bisulphite kit. The Gag region of the IAP and MSR retrotransposons was amplified from the bisulphite-treated genomic DNA using the specific primers. PCR products were cloned and sequenced.

Full Methods and any associated references are available in the online version of the paper.

Received 20 November 2012; accepted 18 July 2013.
Published online 8 September 2013.

1. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16,** 6–21 (2002).

2. Jones, P. A. & Liang, G. Rethinking how DNA methylation patterns are maintained. *Nature Rev. Genet.* **10,** 805–811 (2009).
3. Alabert, C. & Groth, A. Chromatin replication and epigenome maintenance. *Nature Rev. Mol. Cell Biol.* **13,** 153–167 (2012).
4. Margueron, R. & Reinberg, D. Chromatin structure and the inheritance of epigenetic information. *Nature Rev. Genet.* **11,** 285–296 (2010).
5. Probst, A. V., Dunleavy, E. & Almouzni, G. Epigenetic inheritance during the cell cycle. *Nature Rev. Mol. Cell Biol.* **10,** 192–206 (2009).
6. Bostick, M. *et al.* UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317,** 1760–1764 (2007).
7. Sharif, J. *et al.* The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450,** 908–912 (2007).
8. Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y. & Shirakawa, M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* **455,** 818–821 (2008).
9. Avvakumov, G. V. *et al.* Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* **455,** 822–825 (2008).
10. Hashimoto, H. *et al.* The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* **455,** 826–829 (2008).
11. Leonhardt, H., Page, A. W., Weier, H. U. & Bestor, T. H. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71,** 865–873 (1992).
12. Arita, K. *et al.* Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1. *Proc. Natl Acad. Sci. USA* **109,** 12950–12955 (2012).
13. Hu, L., Li, Z., Wang, P., Lin, Y. & Xu, Y. Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2. *Cell Res.* **21,** 1374–1378 (2011).
14. Rajakumara, E. *et al.* PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. *Mol. Cell* **43,** 275–284 (2011).
15. Rothbart, S. B. *et al.* Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nature Struct. Mol. Biol.* **19,** 1155–1160 (2012).
16. Nady, N. *et al.* Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. *J. Biol. Chem.* **286,** 24300–24311 (2011).
17. Song, J., Rechkoblit, O., Bestor, T. H. & Patel, D. J. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **331,** 1036–1040 (2011).
18. Felle, M. *et al.* The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1. *Nucleic Acids Res.* **39,** 8355–8365 (2011).
19. Achour, M. *et al.* The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. *Oncogene* **27,** 2187–2197 (2008).
20. Lutzmann, M. & Mechali, M. MCM9 binds Cdt1 and is required for the assembly of prereplication complexes. *Mol. Cell* **31,** 190–200 (2008).
21. Citterio, E. *et al.* Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Mol. Cell. Biol.* **24,** 2526–2535 (2004).
22. Jenkins, Y. *et al.* Critical role of the ubiquitin ligase activity of UHRF1, a nuclear RING finger protein, in tumor cell growth. *Mol. Biol. Cell* **16,** 5621–5629 (2005).
23. Karagianni, P., Amazit, L., Qin, J. & Wong, J. ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation. *Mol. Cell. Biol.* **28,** 705–717 (2008).
24. Peng, J. *et al.* A proteomics approach to understanding protein ubiquitination. *Nature Biotechnol.* **21,** 921–926 (2003).
25. Kim, W. *et al.* Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* **44,** 325–340 (2011).
26. Shimamura, S. & Ishikawa, F. Interaction between DNMT1 and DNA replication reactions in the SV40 *in vitro* replication system. *Cancer Sci.* **99,** 1960–1966 (2008).
27. Takebayashi, S., Tamura, T., Matsuoka, C. & Okano, M. Major and essential role for the DNA methylation mark in mouse embryogenesis and stable association of DNMT1 with newly replicated regions. *Mol. Cell. Biol.* **27,** 8243–8258 (2007).
28. Plechanovová, A., Jaffray, E. G., Tatham, M. H., Naismith, J. H. & Hay, R. T. Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **489,** 115–120 (2012).
29. Muraki, K., Nabetani, A., Nishiyama, A. & Ishikawa, F. Essential roles of *Xenopus* TRF2 in telomere end protection and replication. *Genes Cells* **16,** 728–739 (2011).
30. Nishiyama, A., Frappier, L. & Mechali, M. MCM-BP regulates unloading of the MCM2–7 helicase in late S phase. *Genes Dev.* **25,** 165–175 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to T. S. Takahashi, K. Shintomi, K. Muraki, H. Nakaoka, M. Iwabuchi and K. Ohsumi for reagents and technical advice, J. Maller for *Xenopus* Orc2 antibody, H. Miyoshi for lentiviral vectors, K. Helin for reading of the manuscript and M. Orii for designing the schematic diagram shown in Supplementary Fig. 1. We thank A. Hosoi and C. Yamada-Namikawa for technical assistance. M.N. was supported by a Grant-in-Aid for Scientific Research on Innovative Areas 'Cell fate control', Scientific Research (A), and Challenging Exploratory Research from MEXT Japan. A.N. was supported by a Research Activity Start-up and Grant-in-Aid for Young Scientists (B) from the Japan Society for the Promotion of Science.

Author Contributions M.N. and A.N. planned studies and interpreted the data. A.N. and L.Y. performed most of the *Xenopus* studies. J.S. and H.K validated the *mUhrf1* knockout lines and performed bisulphite DNA sequencing analysis. Y.J. and K.N. performed most of the mammalian studies. T.Kaw. and T.Kod. performed LC–MS/MS analysis. S.S. and F.I. generated purified recombinant mDnmt1 proteins. K.A. generated the purified recombinant hUhrf1 protein. M.N. and A.N wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.N. (mkt-naka@med.nagoya-cu.ac.jp) or A.N. (anishiya@med.nagoya-cu.ac.jp).

METHODS

Plasmids and proteins. Full-length xUhrf1 was isolated from *Xenopus* total egg cDNA by PCR. The cDNA fragment of xUhrf1 that encodes amino acids 1–650 was cloned into pET16b to be expressed as His-tagged protein. The expression constructs for His-tagged geminin and GST-tagged p27 were gifts from T. S. Takahashi. His- and GST-tagged fusion proteins were expressed in *Escherichia coli* BL21-CodonPlus-RIL cells (Stratagene), purified using standard molecular biology techniques, and stored in aliquots at −80 °C until use. Recombinant His-tagged ubiquitin and its mutant proteins were purchased from Boston Biochem. rmDnmt1 and its mutant were prepared as described previously[26].

Antibodies. Antibodies against xUhrf1 were raised in rabbits by immunization with a His_{10}-tagged, recombinant peptide (1–650) of xUhrf1. The anti-xDnmt1 antibody was raised against a synthetic peptide (CTETVKAEKMETD). Antisera were further affinity-purified with the recombinant protein immobilized onto a nitrocellulose membrane. Anti-xOrc2 was kindly provided by J. Maller. Anti-xPcna (Sigma, PC10), xUsp7 (Bethyl Laboratories, A300-033A), ubiquitin (Santacruz, P4D1), xMcm7 (Abcam, ab2360), H3 (Abcam, ab1791), hDnmt1 (Santacruz, H300), hUhrf1 (BD Biosciences, 612264), mUhrf1 (Santacruz, M-132), HA (MBL, 561), Myc (Santacruz, A-14), Myc (Santacruz, 9E10), Flag (Sigma, M2), His (MBL, PM032), horseradish-peroxidase-conjugated protein A (GE healthcare, NA9120V), (Santacruz, P4D1) and Pcna (Santacruz, PC10) were obtained commercially.

***Xenopus* egg extracts.** Cytostatic factor (CSF)-arrested and interphase egg extracts were prepared as described[29,30], except that metaphase II-arrested eggs were washed in extraction buffer (EB; 5 mM $MgCl_2$, 100 mM KCl, 20 mM HEPES-KOH, pH 7.5) instead of in 1× Marc's modified Ringers (MMR). Incubations of extracts were carried out at 22 °C. For immunoblotting analysis, extracts were diluted 20-fold with SDS sample buffer and boiled for 5 min.

Immunodepletion. For immunodepletion of *Xenopus* egg extracts, 250 µl of anti-serum against xUhrf1 and xDnmt1 were conjugated to 75 µl of protein A agarose (Roche). Twenty-five microlitres of antibody beads were mixed with 100 µl interphase egg extracts and incubated at 4 °C for 1 h. This was repeated three times. For add-back experiments, mDnmt1 wild-type or the C1229S mutant was added to xDnmt1-depleted extracts at 10 ng μl^{-1}.

Monitoring DNA methylation and replication in *Xenopus* egg extracts. DNA methylation was monitored by the incorporation of *S*- [methyl-^{3}H]-adenosyl-L-methionine. Demembranated sperm nuclei (3,000 nuclei per µl) were added to an egg extract containing *S*-[methyl-^{3}H]-adenosyl- L-methionine, incubated at room temperature (22–25 °C), and the reaction was stopped by the addition of chromatin purification buffer (CPB; 50 mM KCl, 5 mM $MgCl_2$, 20 mM HEPES-KOH, pH 7.7, and 2% sucrose) up to 300 µl. His_6–geminin, GST–p27 or aphidicolin (150 µM) was added to the extracts 10 min before the addition of sperm DNA. Genomic DNA was purified using a Wizard Genomic DNA purification kit (Promega) according to the manufacturer's instructions. The DNA replication efficiency was determined as described previously[20].

Isolation of chromatin fractions from *Xenopus* egg extracts. Sperm nuclei were incubated in extracts at a concentration of 3,000 nuclei per µl. Each 30–50-µl extract was diluted fivefold with ice-cold CPB containing 0.1% NP40, overlaid onto a 30% sucrose/CPB cushion, and centrifuged at 15,000*g* for 10 min. The pellets were re-suspended in SDS sample buffer. For MS/MS analysis, to enrich histone proteins, chromatin fractions were washed in 300 mM NaCl and acid-extracted.

Immunoprecipitation. Two micrograms of the antibody for immunoprecipitation were incubated with protein A-agarose beads for 1 h at room temperature and washed in CPB buffer. *Xenopus* egg extracts were diluted fivefold in CPB buffer and incubated with the antibody-bound beads for 1 h at 4 °C. Beads were then washed twice with CPB, once with CPB containing 0.1% Triton X-100, and twice with CPB alone. Proteins were finally eluted with SDS sample buffer and analysed by SDS–PAGE.

Immunoprecipitation of chromatin-bound histone H3. Extracts containing sperm chromatin were diluted fivefold with CPB containing 0.1% NP-40, and were then incubated on ice for 5 min. The chromatin was isolated via centrifugation at 10,000*g* for 10 min through a 0.5-ml sucrose cushion of CPB with 30% sucrose, which was underlayered in the tube with 10 µl of 2 M sucrose. Removal of the supernatant by aspiration left 50 µl and an equal volume of nuclease buffer (micrococcal nuclease) was added. The chromatin pellet was re-suspended by vortexing and incubated at 22 °C for 20 min. The reaction was stopped by the addition of 10 mM EDTA and the tube was centrifuged at 15,000 r.p.m. for 10 min. For H3 immunoprecipitation, the supernatant was supplemented with 1% SDS and was diluted tenfold with lysis buffer (1% Triton X-100, 1 mM EDTA, 150 mM NaCl, 15 mM Tris-HCl, pH 8.0). Diluted lysates were incubated with 2 µg of histone H3 antibody conjugated to 10 µl of protein A agarose. Beads were then washed three times with 400 µl of lysis buffer. For a pulldown experiment, immunoprecipitated histone H3 was incubated with egg extract at 4 °C for 1 h. Beads were washed four times with lysis buffer and bound proteins were finally eluted with SDS sample buffer and analysed by SDS–PAGE.

Pulldown of xUbH3 by immunopurified xDnmt1. Immunoprecipitation of xDnmt1 was performed as described above. Immunopurified xDnmt1 bound to protein A agarose was incubated with 40 µl of the MNase-digested chromatin fraction isolated from either mock- or xDnmt1-depleted extracts. After incubation on ice for 1 h, beads were washed with CPB buffer containing 0.1% NP-40. Bound proteins were analysed by immunoblotting.

Purification of rhUhrf1. A DNA fragment encoding full-length hUhrf1 was amplified by PCR and cloned into pGEX-6P plasmid (GE Healthcare Bio-Sciences). The recombinant protein was overexpressed in the *E. coli* strain Rosetta (DE3). Cells were grown at 37 °C in Luria-Bertani medium, induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) when they reached an optical density of 0.5∼0.6 at 660 nm and were then incubated at 15 °C overnight.

The cells were collected, re-suspended in lysis buffer (50 mM Tris-HCl, pH 8.0, containing 300 mM NaCl, 1 mM dithiothreitol (DTT), 30 µM ZnOAc, 10% glycerol and 0.2 mM phenylmethylsulphonyl fluoride (PMSF)) and disrupted by sonication on ice. After centrifugation, the supernatant was applied to a GST affinity column of glutathione Sepharose 4 Fast Flow (GE Healthcare Bio-Sciences) equilibrated with PMSF-free lysis buffer and eluted by digestion using PreScisson protease. The protein was further purified by anion-exchange chromatography using a HiTrap Q HP column, by DNA binding protein affinity chromatography using a HiTrap Heparin column and by size-exclusion chromatography using a HiLoad 26/60 Superdex200 column (GE Healthcare Bio-Science).

Trypsin digestion and LC–MS/MS. The method for trypsin digestion of protein has been described previously[31]. LC–MS/MS analysis was performed using an LTQ orbitrap ELITE ETD mass spectrometer (Thermo Fisher Scientific). The methods used for LC–MS/MS were slightly modified from those described previously[32]. The mass spectrometer was operated in data-dependent acquisition mode in which MS acquisition with a mass range of *m/z* 350–1,600 was automatically switched to MS/MS acquisition under the control of Xcalibur software. The top three precursor ions in the MS scan were selected by Orbitrap, with a resolution of R = 240,000, and those in subsequent MS/MS scans, with an ion trap in automated gain control (AGC) mode where AGC values were 1×10^6 and 1.00×10^4 for full MS and MS/MS, respectively. For fragmentation, collision-induced dissociation and electron transfer dissociation were used.

Database searching. Tandem mass (MS/MS) spectra were extracted using Proteome Discoverer version 1.3. All MS/MS samples were analysed using Mascot (Matrix Science; version 2.4). Mascot was set up to search Sprot_2012_08.fasta (selected for *Xenopus laevis*, 3,350 entries) assuming the presence of the digestion enzyme trypsin. Mascot was searched with a fragment ion mass tolerance of 0.80 Da and a parent ion tolerance of 3.0 p.p.m. The maximum number of missed cleavage sites was set at 3. The carbamidomethyl of cysteine was specified in Mascot as a fixed modification. Amide of arginine, methyls of lysine and arginine, dimethyls of lysine and arginine, acetyl of lysine, trimethyl of lysine, phosphos of serine, threonine and tyrosine, GlyGly of lysine and LeuArgGlyGly of lysine, all of which have a trypsin-digested ubiquitinated protein signature, were specified in Mascot as variable modifications[24].

Criteria for protein identification. Scaffold (version Scaffold_3.6.3, Proteome Software) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm[33]. Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least three identifiable peptides. Protein probabilities were assigned by the Protein Prophet algorithm[34]. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped together to satisfy the principle of parsimony.

Plasmid construction. To generate lentivirus-based *sh-hDnmt1* or *sh-hUhrf1* constructs, an oligonucleotide containing the nucleotides 303–321 of *hDnmt1* (forward 5′- GATCCCC**GGAGGTGAATGGATGTCTA**ACGTGTGCTGTCC GT**TAGACGTCCATTCACTTCC**TTTTTGGAAAT-3′ and reverse 5′- CTAGA TTTCCAAAAA**GGAAGTGAATGGACGTCTA**ACGGACAGCACACGT**TAG ACATCCACTCACCTCC**GGG-3′; targeting sequence in bold) or nucleotides 1486–1505 of *hUhrf1* (forward 5′-GATCCCC**GGTCGATGGGTACGTTGAT**A CGTGTGCTGTCCGT**ATCGACGTACTCATTGACC**TTTTTGGAAAT-3′ and reverse 5′- CTAGATTTCCAAAAA**GGTCAATGAGTACGTCGAT**ACGGACA GCACACGT**ATCAACGTACCCATCGACC**GGG-3′; targeting sequence in bold) was annealed and inserted into pENTR4-H1 (a gift from H. Miyoshi) digested with AgeI/EcoRI. To insert the H1tetOx1-shRNA into a lentivirus vector, we mixed the resulting pENTR4-H1-shRNA vector and CS-RfA-ETBsd (a gift from H. Miyoshi) or CS-RfA-ETHygro vector with Gateway LR clonase (Invitrogen).

To construct the plasmid, pCAGGS-hH3-Flag, the BamHI (end-filled)/EcoRI (end-filled) fragment containing the full-length cDNA and the Flag epitope from a

pOZ-C-hH3 (a gift of H. Tagami, Nagoya City University) was inserted into a pCAGGS vector digested with EcoRI (end-filled). The hH3(K23R) mutant plasmid was generated by PCR-based, site-directed mutagenesis for pCAGGS-hH3-Flag. To construct pCAG-puro-Myc-mUhrf1, a PCR fragment containing the full-length cDNA was sub-cloned into pCAG-puro-Myc vector. The pCAG-pruo-Myc-mUhrf1 mutant plasmids were generated by PCR-based, site-directed mutagenesis for pCAG-puro-Myc-mUhrf1.

Immunoprecipitation and immunoblotting analyses for mammalian cells. 293T cells or HeLa cells transfected with the indicated constructs were lysed as quickly as possible in TBSN buffer (20 mM Tris-Cl, pH 8.0, 150 mM NaCl, 0.5% NP-40, 5 mM EGTA, 1.5 mM EDTA, 0.5 mM Na3VO4, 50 μM PR-619 and 20 mM *p*-nitrophenylphosphate (PNPP). The resulting lysates were clarified by centrifugation at 15,000*g* for 20 min at 4 °C before immunoprecipitation with the specified antibody. Immunoprecipitated proteins were separated by SDS–PAGE, transferred to a polyvinylidene difluoride (*Immobilon*-P; Millipore) membrane and then detected by immunoblotting with the indicated antibodies using an enhanced chemiluminescence detection system.

Denaturing Immunoprecipitation. Denaturing immunoprecipitation was performed according to the Abcam protocol with a slight modification. In brief, HeLa cells were lysed in denaturing buffer (50 mM Tris-Cl, pH 8.0, 1% SDS, 5 mM EDTA, 10 mM DTT, 15 U ml^{-1} DNaseI, 50 μM PR-619 and protease inhibitors), and the lysates were heated at 95 °C for 10 min and clarified by centrifugation at 15,000*g* for 5 min at 20 °C. The supernatant was diluted ten times with non-denaturing buffer (50 mM Tris-Cl, pH 8.0, 250 mM NaCl, 10% glycerol, 1% NP-40, 5 mM EDTA, 50 μM PR-619 and protease inhibitors), and the lysates were used for immunoprecipitation with Flag M2 agarose. The immunoprecipitates were washed four times with non-denaturing buffer, and used for the binding assay with rmDnmt1 protein.

Cell culture, transfection and virus generation and infection. Cell lines were cultured as recommended by American Type Culture Collection (Manassas). HeLa cells were trapped in S phase by treating them with 2.5 mM thymidine (Sigma) for 20 h. Transfection was routinely carried out with lipofectamine 2000 (Invitrogen). For the production of lentiviruses, transfection was performed by the calcium phosphate co-precipitation method[35]. Lentiviruses expressing shRNA were generated by co-transfecting 293T cells with pCMV-VSV-G-RSV-RevB (a gift from H. Miyoshi), pCAG-HIVgp (a gift of H. Miyoshi), and the respective CS-RfA-ETBsd or CS-RfA-ETHygro vector. HeLa cells infected with the viruses were treated with 10 μg ml^{-1} blasticidin (Invitrogen) and/or 200 ng ml^{-1} hygromycin (Sigma) for 2–3 days. To generate mUhrf1-expressing cells, PvuI-linearized pCAG-puro-Myc-mUhrf1 plasmids were cut with PvuI and transfected into the cells by the calcium phosphate co-precipitation method. Cells transfected with the indicated plasmids were treated with 2 μg ml^{-1} of puromycin (Sigma) for 5 days.

Histone acid extraction. Cell pellets were re-suspended in PBS with 0.5% Triton X-100 and protease inhibitors at a density of approximately 10^7 cells ml^{-1} and tubes were rotated at 4 °C for 10 min to lyse the cells. The lysates were centrifuged at 4 °C for 10 min at 2,000 r.p.m., and the pellets were rinsed once in the extraction buffer. The histones in the insoluble chromatin pellet were then extracted in 0.2 N HCl for 10 min at 4 °C on a rotator. The lysates were centrifuged at 4 °C for 10 min at 2,000 r.p.m., and the supernatants containing histones were collected and adjusted to pH 8.0 with 2 M Tris.

Immunofluorescence analysis. Mouse embryonic fibroblasts grown onto glass slides were treated with 0.1% Triton X-100 in CSK buffer (100 mM NaCl, 300 mM sucrose,10 mM PIPES, pH 6.8, 3 mM $MgCl_2$ and 1 mM EGTA) for 2 min at 4 °C, fixed with 4% paraformaldehyde in PBS, and then treated with methanol for 20 min at −20 °C. The cells were incubated in blocking solution (5% BSA in PBS) and incubated in detection solution containing primary antibodies (5% BSA in PBS). After four washes with PBS, the samples were incubated in detection solution containing the secondary antibodies. Images were collected using BZ-9000 (KEYENCE).

Matintenance of *Uhrf*$^{fl/fl}$ embryonic stem cells. *Uhrf*$^{fl/fl}$ embryonic stem cells were cultured on mitomycin-C-treated non-proliferating feeder mouse embryonic fibroblast cells in standard embryonic stem cell medium supplemented with 20% FBS and LIF. For conditional knockout of the endogenous *mUhrf1* gene, 800 nM 4-OH tamoxifen (4-OHT, SIGMA, no. H7904, dissolved in ethanol) was added to the medium. Ablation of the *mUhrf1* transcript was confirmed by qPCR using the following primers specific for *mUhrf1* exons 4 and 5, mUhrf1-RT-F01, ATG ATTGCAGAGGGGAGTTG; and mUhrf1-RT-R01, GTGGTGAGGGGAGTGA AGAA. Decreases in mUhrf1 protein expression were assessed using an Uhrf1-specific antibody (Uhrf1, M-132 antibody, Santa Cruz Biotechnology, no. sc-98817). Expression of the Myc-tagged exogenous wild-type or mutant Uhrf1 proteins was detected with a mouse monoclonal anti-Myc tag antibody (Millipore, 05-724, clone 4A6).

Wild-type and mutant Uhrf1-expressing mouse embryonic stem cell cloning. Six micrograms of linearized pCAG-IRES-Puro plasmid containing either a Myc-tagged wild-type, an SRA domain mutant (D474A/R489A) or a RING domain mutant (C713A/C715A/C716A) *mUhrf1* cDNA were transfected into *Uhrf1*$^{fl/fl}$ embryonic stem cells using FuGENE HD transfection reagent (Roche Applied Science, no. 6366244). Stable clones were selected by puromycin (1 mg ml^{-1}) treatment. Endogenous *mUhrf1* genes are conditionally deleted by the addition of 4-hydroxytamoxifen from *Uhrf1*$^{fl/fl}$ embryonic stem cells expressing ER-Cre.

Global DNA methylation analysis by methylation-sensitive Southern blotting. Endogenous *mUhrf1* genes were conditionally deleted by the addition of 4-hydroxytamoxifen from *Uhrf1*$^{fl/fl}$ embryonic stem cells expressing ER-Cre. Genomic DNA was purified from embryonic stem cells using an AllPrep DNA/RNA mini kit (QIAGEN, no. 80204) according to the supplier's protocol. One microgram of genomic DNA was treated with 20 U HpaII (New England BioLabs, no. R0171L) or 10 U MspI (New England BioLabs, no. R0106L) overnight at 37 °C. Digests were separated with 0.8% agarose-gel, transferred to a nylon membrane and hybridized with specific probes generated with the following primers: 5′-AGTGGGCAGA GTATTCTCTG-3′; IAP-LTR-probeF, 5′-CAGAAGATTCTGGTCTGTGGTGTT-3′; IAP_LTR_probeR, 5′-GAATTCATACAGTTGAATCCTTCT-3′; MLV_gag_probeF, 5′-TCGAGATCATGGGACAGACC-3′; and MLV_gag_probeR, GGGTAAGGG CAGGGTAAAGG-3′. Relative DNA methylation was determined by counting radioactivity of the undigested DNA (tops of the gels) and represented as a percentage of DNA undigested by HpaII in *Uhrf1*$^{fl/fl}$ mouse embryonic stem cells.

Bisulphite DNA sequencing analysis. Two micrograms of genomic DNA were treated with sodium bisulphite using an EpiTect bisulphite kit (QIAGEN, no. 59104) following the manufacturer's protocol. The Gag region of the IAP retrotransposons was amplified from the bisulphite-treated genomic DNA using the following primers: IAP-Gag-BS-F01, 5′-ATTTTGTTGATTAAATAAATTATTA TTGGG-3′; and IAP-Gag-BS-R01, 5′-TAAAACATATCCTCTAATCATTTCT ACTCA-3′. Primers for minor satellites: forward, 5′-ATTATATTGTAGAATA TATTAGATGAGTGAG-3′; reverse, 5′-ATCATTTTCCATATTTCTCATTAT AACTC-3′. PCR products were cloned into TOPO TA Cloning Dual Promoter vector (Life Technologies, no. K4600-01). Successful clones were sequenced using a BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, no. 4337454). Methylated CpGs are shown as closed circles and unmethylated ones are represented with open circles.

***In vitro* deubiquitylation assays.** For the substrate, hUbH3 was acid-extracted from hDnmt1-depleted HeLa cells which were released 4 h from the G1/S boundary by double thymidine treatment. The hUbH3 was mixed with immune-precipitated xDnmt1 from *Xenopus* egg extract in reaction buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2 mM EDTA, 2 mM DTT) for 8 h at 37 °C. To block the reactions, Laemmli buffer was added to the reaction, and the proteins were analysed by immunoblotting.

31. Fujinoki, M. *et al.* Identification of 36-kDa flagellar phosphoproteins associated with hamster sperm motility. *J. Biochem.* **133,** 361–369 (2003).
32. Daigo, K. *et al.* The proteomic profile of circulating pentraxin 3 (PTX3) complex in sepsis demonstrates the interaction with azurocidin 1 and other components of neutrophil extracellular traps. *Mol. Cell. Proteomics* **11,** M111.015073 (2012).
33. Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74,** 5383–5392 (2002).
34. Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75,** 4646–4658 (2003).
35. Chen, C. & Okayama, H. High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell. Biol.* **7,** 2745–2752 (1987).

doi:10.1038/nature12606

Autophagy promotes primary ciliogenesis by removing OFD1 from centriolar satellites

Zaiming Tang[1,2,3,4], Mary Grace Lin[3], Timothy Richard Stowe[5], She Chen[6], Muyuan Zhu[4], Tim Stearns[5], Brunella Franco[7,8] & Qing Zhong[1,2,3]

The primary cilium is a microtubule-based organelle that functions in sensory and signalling pathways. Defects in ciliogenesis can lead to a group of genetic syndromes known as ciliopathies[1–3]. However, the regulatory mechanisms of primary ciliogenesis in normal and cancer cells are incompletely understood. Here we demonstrate that autophagic degradation of a ciliopathy protein, OFD1 (oral-facial-digital syndrome 1), at centriolar satellites promotes primary cilium biogenesis. Autophagy is a catabolic pathway in which cytosol, damaged organelles and protein aggregates are engulfed in autophagosomes and delivered to lysosomes for destruction[4]. We show that the population of OFD1 at the centriolar satellites is rapidly degraded by autophagy upon serum starvation. In autophagy-deficient *Atg5* or *Atg3* null mouse embryonic fibroblasts, OFD1 accumulates at centriolar satellites, leading to fewer and shorter primary cilia and a defective recruitment of BBS4 (Bardet–Biedl syndrome 4) to cilia. These defects are fully rescued by OFD1 partial knockdown that reduces the population of OFD1 at centriolar satellites. More strikingly, OFD1 depletion at centriolar satellites promotes cilia formation in both cycling cells and transformed breast cancer MCF7 cells that normally do not form cilia. This work reveals that removal of OFD1 by autophagy at centriolar satellites represents a general mechanism to promote ciliogenesis in mammalian cells. These findings define a newly recognized role of autophagy in organelle biogenesis.

During autophagy, the membrane anchored LC3 (microtubule-associated protein 1 light chain 3, also known as MAP1LC3B) interacts with cargo and cargo-adaptor proteins, recruiting cargoes to the autophagosome for subsequent degradation upon fusion of the autophagosome with the lysosome[5–11]. We carried out a tandem-affinity purification using tagged LC3 as bait to search for its interacting proteins (Fig. 1a). In addition to known LC3-interacting proteins (MAP1B, FYCO1, p62 (also known as SQSTM1) and KEAP1[12]), we identified a set of centriolar satellite proteins, including PCM1, OFD1 and CEP131 (also known as AZI1), that had not previously been shown to associate with LC3. PCM1 was also pulled down by LC3 orthologues, GATE16 (also known as GABARAPL2) and GABARAP (Extended Data Fig. 1a). PCM1, OFD1 and LC3 co-immunoprecipitated with each other, indicating that they are in the same complex (Fig. 1b–d). PCM1 probably enhances the interaction between LC3 and OFD1, as the OFD1–LC3 interaction is compromised in PCM1-depleted cells (Fig. 1e, f).

Depletion of PCM1 by RNA interference had no significant effect on autophagy activity as determined by LC3 lipidation and p62 degradation (Extended Data Fig. 1b). We then examined if any of these centriolar satellite proteins is an autophagic substrate. OFD1 protein levels were reduced by serum starvation and this reduction was compromised in autophagy-deficient $Atg5^{-/-}$ mouse embryonic fibroblasts (MEFs) compared to $Atg5^{+/+}$ MEFs, whereas PCM1, IFT88 and BBS4 protein levels were not altered by serum starvation or in $Atg5^{-/-}$ MEFs (Fig. 1g). The messenger RNA levels of OFD1 were not significantly changed upon serum starvation in $Atg5^{+/+}$ and $Atg5^{-/-}$ MEFs (Extended Data Fig. 1c), indicating that OFD1 protein level reduction upon serum starvation is through protein degradation rather than transcriptional downregulation. Blocking autophagic flux by lysosomal inhibitors bafilomycin A1 (Baf) or chloroquine resulted in increased OFD1 accumulation upon serum starvation (Fig. 1h). Taken together, these data indicate that OFD1 is degraded via the autophagy–lysosome pathway upon serum starvation.

OFD1 is the gene underlying the human disease oral-facial-digital syndrome type 1 (OFD1), an X-linked ciliopathy characterized by morphological abnormalities and renal cysts, as well as Joubert syndrome and Simpson–Golabi–Behmel syndrome type 2[13–18]. OFD1 localizes to the distal ends of centrioles and is necessary for distal appendage formation, IFT88 recruitment and primary cilium formation[18,19]. OFD1 also localizes to centriolar satellites, interacting with proteins associated with human ciliary disease, PCM1, CEP290 and BBS4[20]. However, the function of this OFD1 population remains unclear.

The centriolar satellite localization of OFD1 is determined by PCM1, because OFD1 was lost from satellites when PCM1 was depleted (Extended Data Fig. 2a, b). LC3 partially colocalized with PCM1 upon serum starvation in a majority of U_2OS cells expression Myc-LC3, but rarely in unstressed cells (Extended Data Fig. 2c, d). LC3 also partially colocalized with endogenous OFD1 when lysosome activity is blocked by Baf or chloroquine treatment (Extended Data Fig. 3a, b). This colocalization was limited to centriolar satellites, as LC3 did not colocalize with the centriole marker γ-tubulin (Extended Data Fig. 3c, d).

OFD1 was present at both centrioles and centriolar satellites in untreated retinal pigment epithelial (RPE) cells. Remarkably, the centriolar satellite pool of OFD1 was much reduced upon serum starvation, whereas the population of OFD1 at centrioles remained unchanged (Fig. 2a). This serum starvation-induced OFD1 degradation from centriolar satellites was blocked in $Atg5^{+/+}$ MEFs treated with the lysosome inhibitor chloroquine (Extended Data Fig. 4a, b). Notably, PCM1 protein levels are not controlled by autophagy and the centriolar satellite distribution of PCM1 is not altered upon serum starvation (Fig. 1g and Extended Data Fig. 4c–e), indicating that the autophagic degradation is specific to OFD1 at centriolar satellites rather than centriolar satellites as a whole. This notion is further supported by our observation that OFD1 remained at centriolar satellites upon serum starvation in $Atg5^{-/-}$ MEFs but was lost from centriolar satellites in $Atg5^{+/+}$ MEFs (Fig. 2b).

The loss of OFD1 from centriolar satellites was significantly faster than the loss of OFD1 from centrioles, and the rate of loss of OFD1 in $Atg5^{-/-}$ MEFs was slower than that in $Atg5^{+/+}$ MEFs (Extended Data

[1]Center for Autophagy Research, Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. [2]Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. [3]Division of Biochemistry, Biophysics and Structural Biology, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. [4]State Key Laboratory of Plant Physiology and Biochemistry, College of Life Sciences, Zhejiang University, Hangzhou 310058, China. [5]Department of Biology, Stanford University, Stanford, California 94305, USA. [6]National Institute of Biological Sciences, Beijing 102206, China. [7]Medical Genetics Services, Department of Translational Medical Sciences, Federico II University, Via Pietro Castellino 111, Naples, Italy. [8]Telethon Institute of Genetics and Medicine (TIGEM), Via Pietro Castellino 111, Naples, Italy.

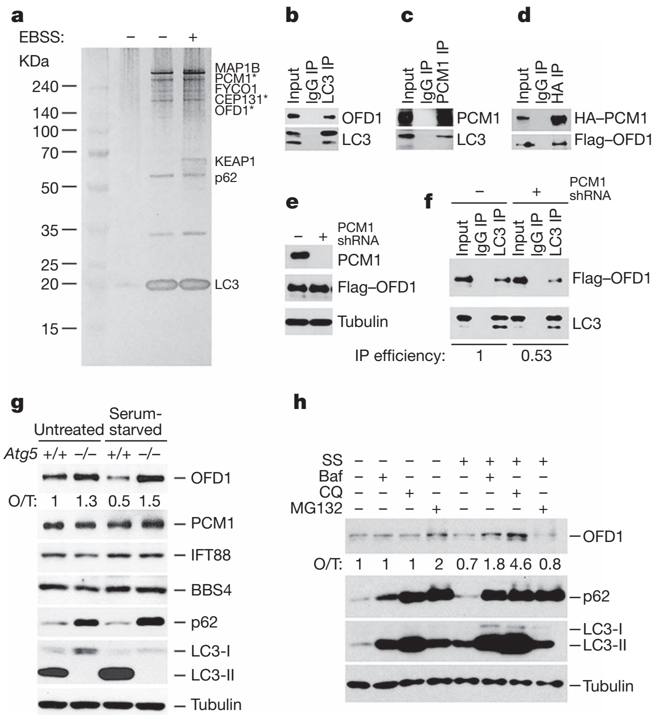


Figure 1 | OFD1 is an autophagic substrate. **a**, Silver staining of LC3 complexes purified from U_2OS cells expressing ZZ–Flag–LC3 in normal medium or Earle's balanced salt solution (EBSS) for 2 h. Asterisks mark centriolar satellite proteins. **b**–**d**, Co-immunoprecipitation (IP) of OFD1 with LC3, LC3 with PCM1, or OFD1 with PCM1 in HEK293T cells. HA, haemagglutinin. **e**, Western blotting analysis of PCM1 and OFD1 protein levels in control or PCM1 knockdown (KD) HEK293T cells. **f**, Co-immunoprecipitation of OFD1 with LC3 in control or PCM1 KD HEK293T cells. IP efficiency is calculated as the ratio of immunoprecipitated OFD1/LC3. **g**, **h**, Western blotting analysis of protein levels of indicated proteins in MEFs with indicated genotypes in normal medium or subjected to 24 h serum starvation (SS), 50 nM bafilomycin A1 (Baf), 20 μM chloroquine (CQ) or 1 μM MG132. Quantified OFD1 level was normalized with β-tubulin. Similar results were observed in three independent experiments.

Fig. 5a). The centriolar satellite pool of OFD1 in *Atg5*$^{+/+}$ cells was lost within 6 h of serum starvation, whereas this pool of OFD1 in *Atg5*$^{-/-}$ cells remained stable even after 24 h of serum starvation (Extended Data Fig. 5b). These data confirm that OFD1 at centriolar satellites has a faster turnover rate than OFD1 at centrioles, and this serum-starvation-induced accelerated degradation is controlled by autophagy.

We next sought to understand how OFD1 regulation might affect ciliogenesis. We observed that the percentage of *Atg5*$^{-/-}$ cells that form a primary cilium, as compared to *Atg5*$^{+/+}$ cells, was significantly reduced and the cilia that did form were shorter (Fig. 3a). This difference was not due to cell cycle regulation (Extended Data Fig. 6a). Lysosome inhibition also compromised primary ciliogenesis in *Atg5*$^{+/+}$ MEFs (Extended Data Fig. 6b–d). The defective ciliogenesis phenotypes were not limited to *Atg5*$^{-/-}$ MEFs, as similar phenotypes were also observed in MEFs lacking another essential autophagy gene, *Atg3* (Extended Data Fig. 6e–g).

We next investigated if the ciliary recruitment of BBS4, a critical event for ciliogenesis[21], is affected in autophagy-deficient cells. We observed that BBS4 accumulated at centriolar satellites in *Atg5*$^{-/-}$ MEFs; more than 50% of cilia had detectable BBS4 in *Atg5*$^{+/+}$ MEFs, whereas only about 10% of cilia were positive for BBS4 in *Atg5*$^{-/-}$ MEFs (Extended Data Fig. 7). These data indicate that BBS4 recruitment to cilia is also defective in autophagy-deficient cells.

If OFD1 accumulation at centriolar satellites is responsible for the ciliary defects in *Atg5*$^{-/-}$ MEFs, we would expect that depletion of OFD1 by RNA interference might rescue these defects. In *Atg5*$^{-/-}$ MEFs stably expressing OFD1 short hairpin RNA (shRNA), with about 50% efficiency of overall depletion of OFD1 (Extended Data Fig. 8a), OFD1 remained at centrioles but was lost from centriolar satellites (Extended Data Fig. 8b-d). Primary cilium formation upon serum starvation was fully restored in these cells; nearly 70% of cells formed a cilium and the length of these cilia was comparable to those formed in *Atg5*$^{+/+}$ MEFs (Fig. 3b). Hence, the centriolar satellite pool of OFD1 indeed plays a key role in suppressing primary ciliogenesis in *Atg5*$^{-/-}$ MEFs. Notably, with OFD1 knockdown, nearly 45% of the *Atg5*$^{-/-}$ cells formed cilia even without serum starvation (Fig. 3b), indicating that OFD1 degradation is likely to be required for serum-starvation-induced primary ciliogenesis. This notion is supported by

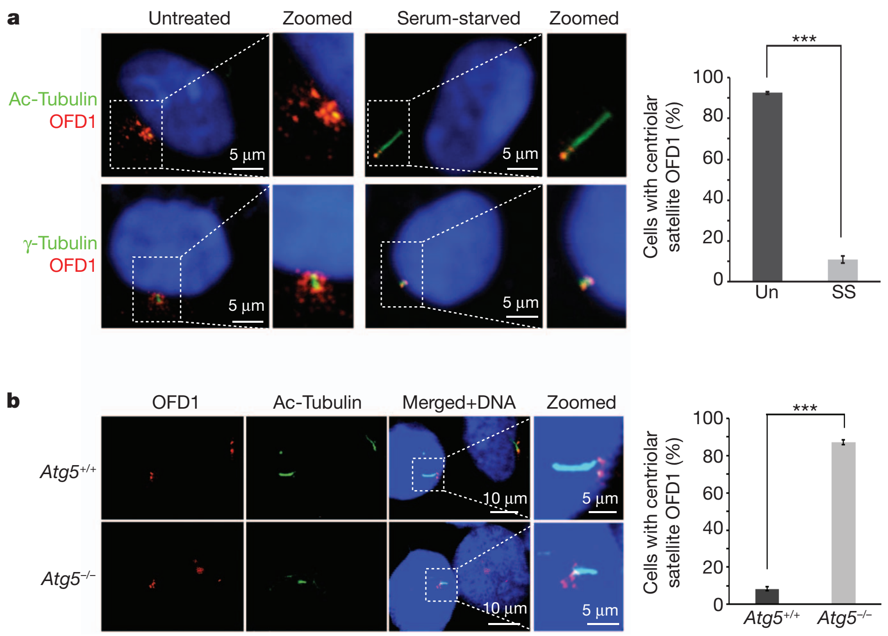


Figure 2 | Autophagy specifically degrades OFD1 at centriolar satellites upon serum starvation. **a**, Representative confocal images of endogenous OFD1 puncta with cilium marker acetylated tubulin (Ac-tubulin), or centriole marker γ-tubulin in hTERT-RPE1 cells in normal medium (Un) or subjected to 24 h serum starvation (SS). **b**, Representative confocal images of endogenous OFD1 and acetylated tubulin in *Atg5*$^{+/+}$ and *Atg5*$^{-/-}$ MEFs subjected to 24 h serum starvation. **a**, **b**, Data shown represent mean ± s.d. percentage of cells with centriolar satellite OFD1 for 100 cells per well in triplicate samples. ***$P < 0.001$, two-tailed unpaired student's t-test. Similar results were observed in three independent experiments.

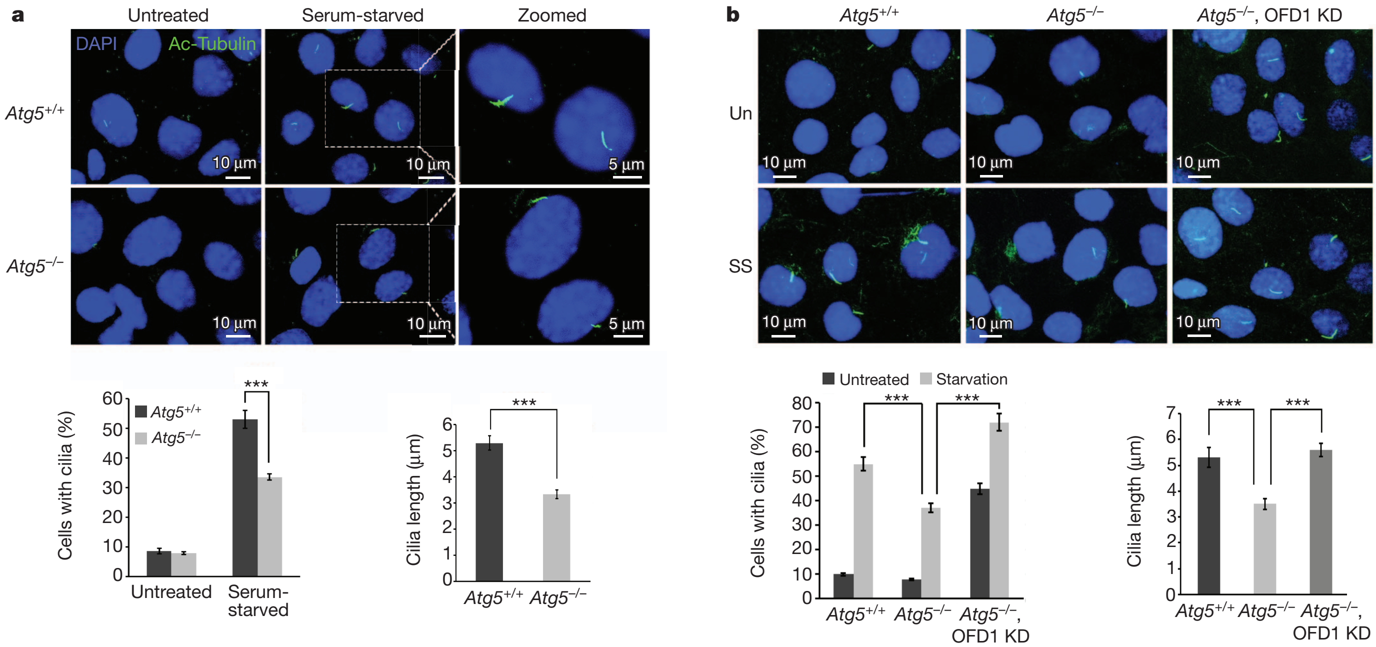


Figure 3 | Autophagy promotes primary cilia biogenesis by regulating OFD1 levels. **a**, **b**, Representative confocal images of cilium marker acetylated tubulin from MEFs with indicated genotype subjected to 24 h serum starvation. Data shown represent mean ± s.d. percentage of cells with primary cilia or length of the cilia for 500 cells or 100 cells per well, respectively, in triplicate samples. $***P < 0.001$, two-tailed unpaired student's *t*-test. Similar results were observed in three independent experiments. Un, untreated; SS, serum-starved.

the observation that primary cilia formed efficiently even in the cycling $Atg5^{+/+}$ MEFs depleted of OFD1 without serum starvation (Extended Data Fig. 9a). The presence of cilia in MEFs was further confirmed by the positive staining of the cilium marker ARL13B (Extended Data Fig. 9b), which specifically decorates the ciliary membrane[22].

Primary cilia are formed in normal breast epithelial cells, but not detected in breast cancer cell lines such as MCF7[23]. Restoration of primary cilia in human cancer cells might be beneficial to reduce malignancy, because cilia are required for proper functions of several signalling pathways (hedgehog etc.) that have crucial roles in many types of cancers[24]. We further tested if OFD1 reduction could restore primary ciliogenesis in MCF7 cells. We generated an OFD1 stable knockdown clone (C19) in MCF7 cells, which showed 44% OFD1 depletion efficiency (Fig. 4a). Similar to OFD1 knockdown MEFs, the centriolar satellite population of OFD1 in C19 cells was markedly reduced, whereas the centriole pool remained stable (Extended Data Fig. 10a–c). Remarkably, primary cilia formed in about 20% C19 cells upon serum starvation, but were completely absent in parental MCF7 cells (Fig. 4b). The length of these cilia ranged from 2 to 6 μm (Fig. 4c,d). As expected for normal primary cilia, these cilia extended from basal bodies and were positive for ARL13B, acetylated-tubulin and the intraflagellar transport protein IFT88 (Fig. 4e and Extended Data Fig. 10d–f). Finally these cilia could be directly visualized under the scanning electron microscope (Fig. 4f). Thus, by manipulating the expression of one protein, OFD1, it is possible to reverse the cilia-defective phenotype of a transformed breast cancer cell line. Primary

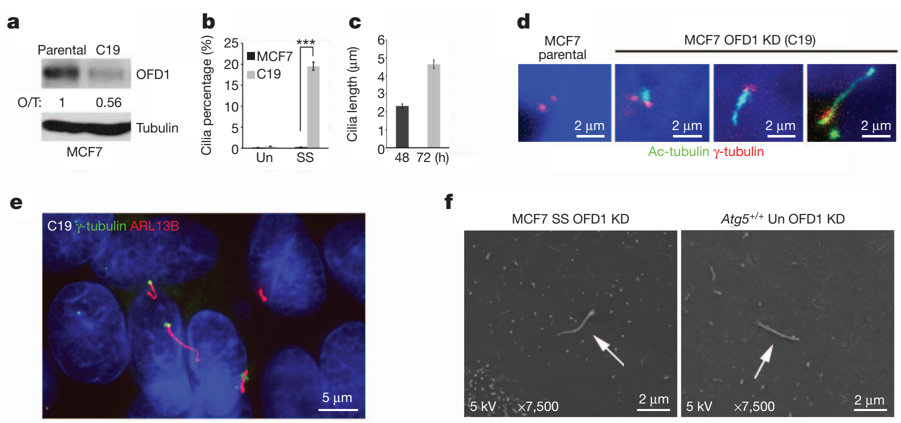


Figure 4 | Forced OFD1 reduction promotes ciliogenesis in human breast cancer MCF7 cells. **a**, Western blotting analysis of OFD1 protein levels in control or OFD1 knockdown MCF7 (C-19) cells, quantified OFD1 level was normalized with β-tubulin. **b**, Quantification of percentage of cells with primary cilia in MCF7 or C19 cells subjected to 72 h serum starvation. Data shown represent mean ± s.d. percentage of cells with primary cilia for 500 cells per well in triplicate samples. $***P < 0.001$, two-tailed unpaired student's *t*-test. **c**, Quantification of length of primary cilia in C19 cells subject to 48 h or 72 h serum starvation. **d**, **e**, Representative confocal images of primary cilia with variable length formed in MCF7 or C19 cells subject to 72 h serum starvation. Data shown represent 100 cells per well in triplicate samples. **f**, Scanning electron microscope analysis of primary cilia (marked by arrows) formed in C-19 cells subject to 72 h serum starvation and cycling $Atg5^{+/+}$ MEFs in normal medium for 24 h. Similar results were observed in three independent experiments.

ciliogenesis remains inefficient in C19 cells, indicating that addition factors are required to reach full capacity. Nevertheless, these data confirm that OFD1 at centriolar satellites functions as a crucial suppressor of primary ciliogenesis in human cancer cells.

We show here that autophagy is linked to primary ciliogenesis, by controlling the degradation of OFD1 at centriolar satellites. We propose that autophagy deficiency serves as a potential underlying mechanism of ciliopathies, and that autophagy modulation might provide a novel means of ciliopathy treatment. We demonstrate that OFD1 at centriolar satellites has a crucial role in suppressing primary ciliogenesis, whereas OFD1 at centrioles has been shown to be essential for primary ciliogenesis. Removing OFD1 from centriolar satellites promotes ciliogenesis in autophagy-deficient cells, wild-type MEFs without serum starvation, and human breast cancer MCF7 cells that normally completely lack cilia, suggesting a general role of OFD1 in suppressing ciliogenesis. Primary ciliogenesis is defective in human breast and pancreatic cancer cells, but activated in the corresponding normal tissues/cells[23,25]. The contribution of primary cilium function to tumorigenesis is complex[26,27]; however, our results suggest that dissecting the regulatory mechanisms of OFD1 will provide insight into these functions and potentially offer new therapeutic tools for treatment of ciliopathies and cancers.

METHODS SUMMARY

Primary cilia biogenesis assay. Cells were incubated in either fresh normal medium (10% FBS, 1% P/S DMEM) or serum-starved medium (0.5% FBS, 1% P/S DMEM) for 24 h. Cells were seeded at different numbers to ensure that they reached the same confluency 1 day after treatment.

Tandem affinity purification of LC3 complexes. This protocol was adapted from previously described methods[12,28,29]. Cell lysates from inducible cell lines expressing ZZ–Flag–LC3 treated with 10 ng ml^{-1} doxycycline for one day were purified through immunoglobulin G (IgG) beads followed by tobacco etch virus (TEV) protease cleavage and anti-Flag M2 affinity gel. All LC3-interacting proteins were identified by mass spectrometry analysis.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 October 2012; accepted 29 August 2013.

Published online 2 October 2013.

1. Sang, L. *et al.* Mapping the NPHP-JBTS-MKS protein network reveals ciliopathy disease genes and pathways. *Cell* **145,** 513–528 (2011).
2. Hildebrandt, F., Benzing, T. & Katsanis, N. Ciliopathies. *N. Engl. J. Med.* **364,** 1533–1543 (2011).
3. Nigg, E. A. & Raff, J. W. Centrioles, centrosomes, and cilia in health and disease. *Cell* **139,** 663–678 (2009).
4. Levine, B. & Klionsky, D. J. Development by self-digestion: molecular mechanisms and biological functions of autophagy. *Dev. Cell* **6,** 463–477 (2004).
5. Thurston, T. L., Ryzhakov, G., Bloor, S., von Muhlinen, N. & Randow, F. The TBK1 adaptor and autophagy receptor NDP52 restricts the proliferation of ubiquitin-coated bacteria. *Nature Immunol.* **10,** 1215–1221 (2009).
6. Johansen, T. & Lamark, T. Selective autophagy mediated by autophagic adapter proteins. *Autophagy* **7,** 279–296 (2011).
7. Kirkin, V., McEwan, D. G., Novak, I. & Dikic, I. A role for ubiquitin in selective autophagy. *Mol. Cell* **34,** 259–269 (2009).
8. Novak, I. *et al.* Nix is a selective autophagy receptor for mitochondrial clearance. *EMBO Rep.* **11,** 45–51 (2010).
9. Orvedahl, A. *et al.* Image-based genome-wide siRNA screen identifies selective autophagy factors. *Nature* **480,** 113–117 (2011).
10. Zhang, Y. *et al.* SEPA-1 mediates the specific recognition and degradation of P granule components by autophagy in *C. elegans*. *Cell* **136,** 308–321 (2009).
11. Kirkin, V. *et al.* A role for NBR1 in autophagosomal degradation of ubiquitinated substrates. *Mol. Cell* **33,** 505–516 (2009).
12. Fan, W. *et al.* Keap1 facilitates p62-mediated ubiquitin aggregate clearance via autophagy. *Autophagy* **6,** 614–621 (2010).
13. Feather, S. A., Woolf, A. S., Donnai, D., Malcolm, S. & Winter, R. M. The oral-facial-digital syndrome type 1 (OFD1), a cause of polycystic kidney disease and associated malformations, maps to Xp22.2-Xp22.3. *Hum. Mol. Genet.* **6,** 1163–1167 (1997).
14. Coene, K. L. *et al.* OFD1 is mutated in X-linked Joubert syndrome and interacts with LCA5-encoded lebercilin. *Am. J. Hum. Genet.* **85,** 465–481 (2009).
15. Field, M. *et al.* Expanding the molecular basis and phenotypic spectrum of X-linked Joubert syndrome associated with *OFD1* mutations. *Eur. J. Hum. Genet.* **20,** 806–809 (2012).
16. Zullo, A. *et al.* Kidney-specific inactivation of Ofd1 leads to renal cystic disease associated with upregulation of the mTOR pathway. *Hum. Mol. Genet.* **19,** 2792–2803 (2010).
17. Ferrante, M. I. *et al.* Identification of the gene for oral-facial-digital type I syndrome. *Am. J. Hum. Genet.* **68,** 569–576 (2001).
18. Ferrante, M. I. *et al.* Oral-facial-digital type I protein is required for primary cilia formation and left-right axis specification. *Nature Genet.* **38,** 112–117 (2006).
19. Singla, V., Romaguera-Ros, M., Garcia-Verdugo, J. M. & Reiter, J. F. *Ofd1*, a human disease gene, regulates the length and distal structure of centrioles. *Dev. Cell* **18,** 410–424 (2010).
20. Lopes, C. A. *et al.* Centriolar satellites are assembly points for proteins implicated in human ciliopathies, including oral-facial-digital syndrome 1. *J. Cell Sci.* **124,** 600–612 (2011).
21. Nachury, M. V. *et al.* A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. *Cell* **129,** 1201–1213 (2007).
22. Cantagrel, V. *et al.* Mutations in the cilia gene *ARL13B* lead to the classical form of Joubert syndrome. *Am. J. Hum. Genet.* **83,** 170–179 (2008).
23. Yuan, K. *et al.* Primary cilia are decreased in breast cancer: analysis of a collection of human breast cancer cell lines and tissues. *J. Histochem. Cytochem.* **58,** 857–870 (2010).
24. Goetz, S. C. & Anderson, K. V. The primary cilium: a signalling centre during vertebrate development. *Nature Rev. Genet.* **11,** 331–344 (2010).
25. Seeley, E. S., Carriere, C., Goetze, T., Longnecker, D. S. & Korc, M. Pancreatic cancer and precursor pancreatic intraepithelial neoplasia lesions are devoid of primary cilia. *Cancer Res.* **69,** 422–430 (2009).
26. Wong, S. Y. *et al.* Primary cilia can both mediate and suppress Hedgehog pathway-dependent tumorigenesis. *Nature Med.* **15,** 1055–1061 (2009).
27. Han, Y. G. *et al.* Dual and opposing roles of primary cilia in medulloblastoma development. *Nature Med.* **15,** 1062–1065 (2009).
28. Sun, Q. *et al.* Identification of Barkor as a mammalian autophagy-specific factor for Beclin 1 and class III phosphatidylinositol 3-kinase. *Proc. Natl Acad. Sci. USA* **105,** 19211–19216 (2008).
29. Chen, D. *et al.* A mammalian autophagosome maturation mechanism mediated by TECPR1 and the Atg12-Atg5 conjugate. *Mol. Cell* **45,** 629–641 (2012).

Acknowledgements We thank D. Moughon and W. Fan for initial purification of LC3 complex; N. Mizushima for *Atg5*$^{+/+}$ and *Atg5*$^{-/-}$ MEFs and M. Komatsu for *Atg3*$^{+/+}$ and *Atg3*$^{-/-}$ MEFs; H. Zhou for helpful discussions and quantitative RT–PCR data analysis; A. Kodani and J. F. Reiter for reagents, helpful discussions and technical assistance; and B. Levine for helpful discussions and critical reading of the manuscript. The electron microscopy studies were performed by C. Miller and V. Gattone II at the Indiana University School of Medicine Electron Microscopy Center that is supported by the Polycystic Kidney Disease Foundation. This work was supported by grants from the American Cancer Society (RSG-11-274-01-CCG) and National Cancer Institute (CA133228) to Q.Z. and grants from the Italian Telethon Foundation (TGM11CB3) and the EU FP7/2007-2013 Programme (Syscilia grant agreement no. 241955) to B.F. The work was partially supported by China Scholarship Council to Z.T.

Author Contributions Z.T., M.G.L. and T.R.S. performed the experiments; S.C. carried out mass spectrometry analysis; M.Z., T.S. and B.F. provided technical and intellectual support; Z.T. and Q.Z. conceived the project, designed the experiments, analysed the data and wrote the manuscript with the help of all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Q.Z. (qing.zhong@utsouthwestern.edu).

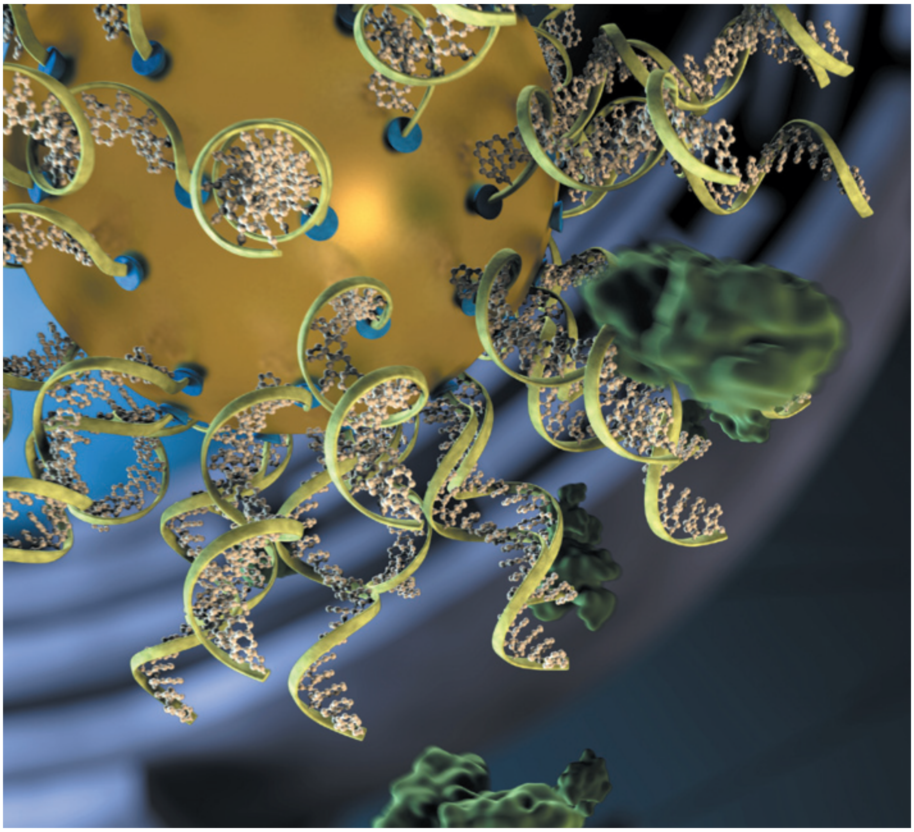
CAREERS

TURNING POINT Genomicist navigates Spain's difficult funding climate **p.261**

@NATUREJOBS Follow us on Twitter for the latest news and features **go.nature.com/e492gf**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

Q. ANDERSON/C. MIRKIN



A computer animation illustrates the activity of particles called spherical nucleic acids.

COLUMN

Animated science

Understanding how visualizations can communicate research will help scientists to make the most of the technology, says **Quintin Anderson**.

As an animator, I help scientists to diffuse their ideas throughout the research community, raise awareness of their work and attract funding. But during my most recent project, I realized that many scientists could benefit from more information so that they can use animation to its full potential.

My client was Chad Mirkin, director of the International Institute for Nanotechnology at Northwestern University in Evanston, Illinois. He wanted an animation that whisked the viewer to the molecular level for a look at the properties of spherical nucleic acids (SNAs) — nanoparticles developed and researched at his lab. But it was his first time producing an animation. He wanted to know what his role would be. How would we keep the audience's attention? How would we communicate the science accurately and clearly?

Some basic tips will help to demystify science animation, and will let scientists take full advantage of this powerful tool.

The first step is to write a script describing the concepts and ideas that the researcher hopes to illustrate and communicate. The researcher can start by setting out a synopsis of the science, working alone or with the animator. Then he or she can break the synopsis down into scenes, and use it as a guide to write the narration. For example:

Scene 1 synopsis: SNAs are defined.
Narrator: Spherical nucleic acids, or SNAs, consist of densely packed, highly oriented nucleic acids, typically arranged on the surface of a nanoparticle.

Narration should be concise. The overall length of animations can vary, but generally a video explaining a research idea or hypothesis will last for two to four minutes. A narrator will typically speak 125–130 words per minute, so a three-minute animation will contain no more than about 400 words of narration. Depending on the lag time between revisions, it can take two to four weeks to complete a script for a three-minute animation.

PROVIDE CONTEXT

Narration should mention the benefits or impact of the science so that the audience can relate to the content. Consider, for example, a hypothetical animation on insulin and the insulin receptor. A script simply describing the relationship between these molecules, without mentioning why their interaction is important, might read:

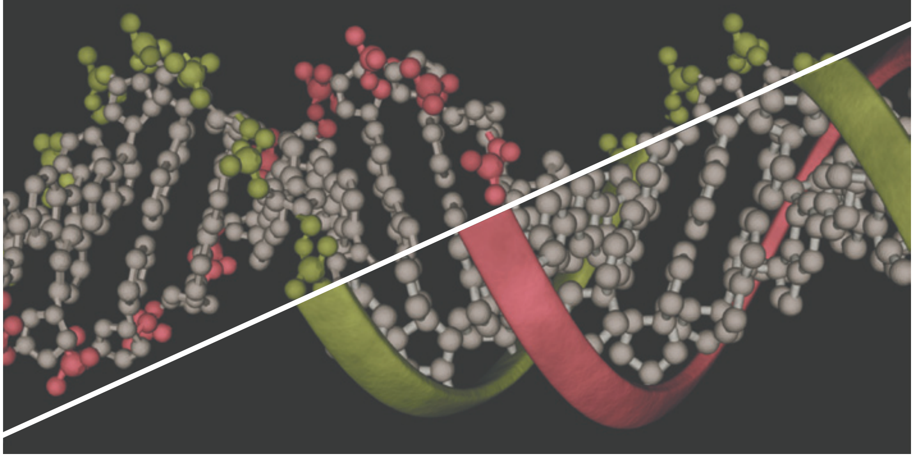
Scene 1 synopsis: The relationship between insulin and the insulin receptor is described.
Narrator: The binding of insulin to the insulin receptor leads to an intracellular signalling cascade that promotes the storage of glucose in cells. The failure of insulin to bind to the insulin receptor can prevent the cellular uptake of glucose.

To be more engaging, however, the narration could describe how the molecules' interactions affect health:

Scene 1 synopsis: The relationship between insulin and the insulin receptor is examined in the context of diabetes.
Narrator: The binding of insulin to the insulin receptor leads to an intracellular signalling cascade that promotes the storage of glucose in cells. If insulin fails to bind to the insulin receptor, the cells don't take up glucose, and blood sugar spikes, producing symptoms commonly associated with diabetes.

VISUAL APPROACH

The ribbon model, used for the helix on the right, illustrates the shape and contour of the molecule. The ball-and-stick model, used for the helix on the left, emphasizes the molecule's atoms and the bonds between them. Which model is best depends on what the researcher wishes to highlight.



Q. ANDERSON

As with any communication about science, the audience's scientific background and level of education must be taken into account. Mirkin's animation was aimed at viewers with an undergraduate-level understanding of chemistry and biology. Mirkin, therefore, used biological language to describe how SNAs enter cells: "Their high density of nucleic acids leads to interactions with membrane-bound scavenger proteins. As a result they are taken up through endocytosis."

But in an interview with The Good Stuff (thegoodstuffshow.com), a website that develops informational videos on scientific phenomena for the layperson, Mirkin used nontechnical language: "You can feed them directly into cells, and the cells just soak them up." Pinpointing the target audience before writing the script is crucial.

MODEL BEHAVIOUR

After completing the script, the scientist works with the animator to turn his or her vision into reality. The chances of a productive interaction and collaboration are greatly increased if the researcher understands how aspects of animation such as models and colours can be used to communicate science effectively.

The animator builds models to visualize events. Different types of model help to evoke different aspects of the data or features of the players involved. Ribbon models, for example, are effective for illustrating the shape and contour of a large molecule. Ball-and-stick models emphasize the molecule's atoms and the bonds between them. Which model is most appropriate depends on what properties the researcher wishes to highlight (see 'Visual approach').

In Mirkin's animation, for example, the design of short strands of DNA and RNA — oligonucleotides — required both ribbon and ball-and-stick models. To make the oligonucleotides easy to identify, Mirkin wanted their backbones to look like a helix — a shape commonly associated with nucleic acids. For that, we used a ribbon model.

However, the ribbon model was not appropriate for representing the bases connected to the backbone. It simplified each to a flat plank, making cytosine, for instance, indistinguishable from adenine. To differentiate the bases, we represented their atoms as balls, and the bonds between these atoms as sticks.

In other cases, where the emphasis is not on the chemical bonds and atoms, but on the space that the atoms occupy, I represent the atoms using 'space-filling' models, in which each atom is shown as a sphere that butts up against the next. This depicts the molecule as a solid object with an uneven surface.

KEEP IT SIMPLE

I urge my clients to remember that models simplify the underlying data. If the model does not capture the science adequately, the researcher should brainstorm with the animator to make it more suitable.

> ***"Animation can be not only fun, but also incredibly effective."***

Although the shape and contour possibilities are many, they are not endless. Software can create a variety of models, but each one takes up computer resources, so scenes with thousands of objects are difficult to show, especially if those models are very detailed. Lipid bilayers, for example, are notoriously hard to craft because they contain hundreds of thousands of lipid molecules.

Scientific accuracy sometimes has to be compromised. For example, the space inside cells is crowded with proteins. If this were shown, depicting a single event in the cell would be difficult. Therefore, this space is commonly left empty in animations, to demonstrate more clearly and cleanly the biochemical events that are the focus of the narrative. Brownian motion — the random movement of particles caused by molecular collisions — may have to be omitted or reduced if it impedes the view of the science. Simplified representations might not be completely accurate, but they help to communicate the science clearly in a short time.

ARTISTIC IMPRESSION

Colour and sound can reinforce the scientific details. For example, adopting a meaningful colour scheme can help the viewer to make sense of a complex scene. In a cross-section of the bacterium *Escherichia coli*, David Goodsell, a molecular-illustration pioneer at the Scripps Research Institute in La Jolla, California, colour-coded molecules to identify the different functional compartments of the cell. But proceed with caution to avoid confusion. If, for example, the cytoplasmic proteins in the *E. coli* cross-section had been given the same colour as the membrane-bound proteins, then these compartments would have been hard to tell apart.

Adding music can make an animation more engaging. BioVisions, a multimedia lab at Harvard University in Cambridge, Massachusetts, accompanied an animation called *The Inner Life of the Cell* with an evocative piano piece. In my view, the music was not just for entertainment. The repetitive melody mirrored the rhythmic cycles of cellular processes, enriching the feeling that these processes were both understandable and eternal. The crescendos enhanced the beauty of these basic scientific processes. The music, which won the 2006 Telly Award for Best Music Composition for a Non-Broadcast Film or Video, helped the viewer to become immersed in the intracellular world, and helped to create a lasting impression.

But sometimes music detracts from the aim of the animation. Again, knowing one's audience is key. BioVisions made a version of *The Inner Life of the Cell* in which a narrator describes the intracellular processes illustrated. It had no music, and for good reason: a score could have become distracting for, say, a student using the animation to prepare for a cell-biology exam. Scientist and animator should determine the objectives and audience of the animation to work out whether music helps or hinders.

Animation can be not only fun, but also incredibly effective. By crafting an engaging, audience-appropriate script, and providing proper feedback to the animator, scientists can turn their research into something more than facts, figures and acronyms. They can craft it into a compelling visual story. ■

Quintin Anderson *is the creative director of the Seagull Company in Midland, Texas.*

CORRECTION

The Careers Feature 'Financial Burden' (*Nature* **501,** 579–581; 2013) wrongly stated that TIAA-CREF is based in Charlotte, North Carolina. It is based in New York City.

TURNING POINT
Ben Lehner

As an early-career scientist, Ben Lehner got swept up in the explosion of genomics-related research. Now at the Centre for Genomic Regulation (CRG) in Barcelona, Spain, working with the European Molecular Biology Laboratory (EMBL), he won the 2013 Eppendorf Young Investigator Award for teasing apart the relationships between genotypes and phenotypes. Despite fragile funding, he finds the Barcelona science scene to be vibrant.

You trained in Cambridge, UK. What was that like in the early genomics era?
There was a can-do attitude. It felt like a historic time. During my last undergraduate year at the University of Cambridge, the first draft of the human genome sequence was released. I joined a genomics lab, and worked on mapping human protein interactions. But I got frustrated — it's difficult to cleanly manipulate human genes.

You moved on to the nematode *Caenorhabditis elegans*. What research did that allow?
The *C. elegans* community had built a genome-wide RNA interference library to knock down any gene. It opened my eyes; it was much easier to do experiments. My postdoc lab at the Wellcome Trust Sanger Institute near Cambridge, systematically looked at genetic interactions. This crazy, manual screening experiment was set up as a factory and took several months.

You next applied to the CRG. What was the interview process like?
I was used to Anglo-American-style interviews, which are one-on-one, but this was more like the German system, with many candidates in front of a big panel. I didn't get particularly good feedback and didn't hear anything for a few weeks — until the offer.

How did you know it would be a good move?
When I went to the CRG and saw people moving from the United States and Germany, I knew it was a unique, world-class institute in a city of creative people — a place that I would want to live in. And it has a non-tenure-track system so junior groups are not competing. But investigators are supposed to stay for nine years at most.

When is your time up?
Officially, very soon. I have been here seven years now. Unofficially, I want to stay.

How might you make a case for staying?
I will say that I want to stay if it's a possibility. My group's work is central to the interests of this institute, so that is in my favour. But at the end of the day, it will come down to whether they have space for another senior person. In general, I support a non-tenured system because it brings in new ideas. But it would be a disaster if every institute tried to follow this model.

You have described Barcelona's funding as strong but fragile. How so?
The regional government of Catalonia has done a much better job of protecting research budgets than the national government during the economic downturn, and this regional funding is the main source of income for institutes such as the CRG. Science funding from Spain's government has gone down in recent years, but the regional government is protecting it as best it can.

Is your funding secure?
At the CRG, in theory, you do not have to rely on external grants. But there is pressure to bring in funds if you want a larger lab. However, even if I keep getting money, I can't keep expanding. The institute's goal is not to grow empires but to have more groups that do a few things well.

Will you be looking for US opportunities?
I'm not sure. The United States is not particularly attractive right now because the funding success rate is so low. My friends there are having a difficult time. Lots of people are moving back to Europe or to China. But it has to be kept in perspective. The United States still has enormous numbers of fantastically funded groups.

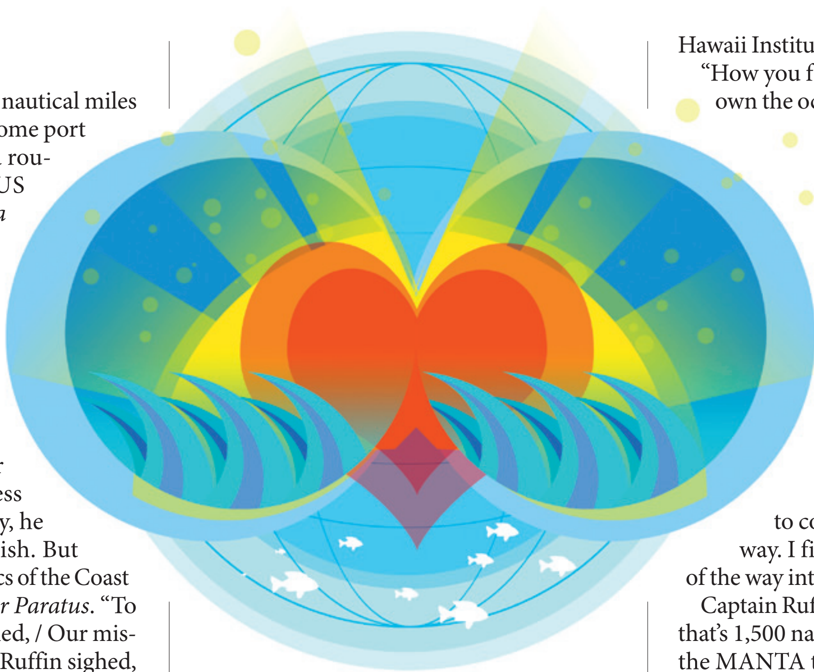
How do you make the big career decisions?
I find that the really important decisions in life are never made in a rational way. They just feel right and you go with it. It is the smaller decisions that get made rationally. ■

INTERVIEW BY VIRGINIA GEWIN

WAVEHITCHER

The course of true love.

BY PAUL DI FILIPPO


JACEY

At a point roughly 100 nautical miles northeast out of its home port of Hilo, Hawaii, on a routine fisheries patrol, the US Coast Guard Cutter *Kiska* encountered a wavehitcher.

Standing on the bridge and peering through his digital recording binoculars, Captain Don Ruffin swore, heedless of how his intemperance would look on the official record. This idiot of a wavehitcher was going to cost him endless paperwork and delay. Briefly, he debated letting the fool perish. But then, unbidden, came the lyrics of the Coast Guard's service song, *Semper Paratus*. "To sink the foe or save the maimed, / Our mission and our pride." Captain Ruffin sighed, put down his binocs, and addressed Lieutenant Commander Billie Schafer with more wistfulness than hope.

"Have you pinged that AUV yet, Lieutenant? Maybe it's authorized to carry a technician or a scientist on a short-term basis."

"Negative, sir. It's a Liquid Robotics MANTA model, licensed to the C-More group at the Hawaii Institute of Marine Biology. Strictly an autonomous underwater vehicle. Tasked to conduct a survey of microbial ecology."

"All right then. We've got an illegal rider. Let's haul him in." Submitting to a sudden impulse, Captain Ruffin added, "I'm going on the pick-up ride, Lieutenant Commander. You have the *Kiska*."

At the starboard side, seamen were unlimbering the cutter's rigid-hull inflatable and getting it launched. Wearing his smart lifevest, Captain Ruffin was soon aboard and skimming across the warm green waters.

The wavehitcher must have sensed their approach, through vibrations alone. But there was no place he could flee, no action he could take. So he just continued to trail the slowly moving AUV at the length of the tether he had secured to its stubby mast.

The MANTA craft was a sleek hull bristling with sensors, solar panels and communications devices. Unseen, well below the surface, was the attached set of vanes that propelled the AUV simply by taking advantage of the constant up-and-down movements of the craft.

Captain Ruffin had been prepared to see the wavehitcher clad in a typical joyrider's amateur rig. But no, this wavehitcher sported very high-tech gear. Could he be a pirate — or a terrorist even? What could be his objective, out here in the middle of nowhere? He wasn't even in a registered shipping lane, and the MANTA was no doubt programmed to avoid such busy corridors. Still, Captain Ruffin could take no chances, so he ordered one of his men to train a weapon on the trespasser.

Soon they had grappled the man aboard. He put up no resistance, and the only difficulty arose in dealing with a capacious seine of some sort that had been attached to him, trailing unseen beneath the sea.

Sitting under armed guard, the wavehitcher exhibited no nervous body language in his superhero wetsuit, which covered even his head. Instead, he slowly lifted his hands up to remove his goggles and breathing apparatus, the latter just a simple snorkel. Friendly blue eyes in a tanned surfer's face and a big ingenuous smile greeted his captors.

"Hey, sapes, what's filtering? You need my help with something?"

Captain Ruffin stared incredulously. "What is your name?"

"Andy Weeden. But you can call me 'Weed'. Everybody does."

"Mr Weeden, are you aware that you were trespassing on the property of the Hawaii Institute of Marine Biology?"

"How you figure that, sapes? They don't own the ocean. Nobody does."

"But this is their registered drone, legal property of the institute."

"I know that. I was just drifting friendly like alongside."

"Nonsense! You were tethered to its mast!"

"All right, all right, so I was hitching a ride. I wasn't hurting it none."

"And exactly where did you think you were heading?"

"This probe is programmed to come within ten miles of Midway. I figured on swimming the rest of the way into shore."

Captain Ruffin's jaw dropped. "But — but that's 1,500 nautical miles away! At the rate the MANTA travels, it would take you six weeks to get there!"

"No problem, sapes! My goggles are smart, my earbuds high-quality and my haptic controls totally sizzling. The Internet feed I tapped from the MANTA was zealous! Besides, I had a lot of thinking to do. Six weeks would have been just about right. You see, my girlfriend is on Midway. We had a scandonkulous fight, and she left before I could apologize and make up. She works for the Fish and Wildlife Service on the Atoll, and they've interdicted all civilians, 'cuz of that stupid Chinese situation. This was the only way I could get to see her."

"How did you intend to survive?"

Weed was happy to explain. His smartsuit used nanoporous graphene membranes to desalinate plenty of drinking water. An array of pulsed magnets repelled sharks. And his seine caught krill and processed it into a nutritious paste routed into a CamelBak reservoir.

"Just call me a one-man cruise ship!" Weed boasted.

The inflatable pulled up next to the *Kiska*. Weed looked forlorn.

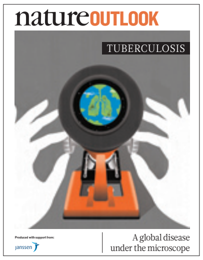
"C'mon, sapes, you're not gonna stand in the path of true love, are you?"

Captain Ruffin looked at his grinning seamen, then said: "Throw him back, he's undersized." ■

Paul Di Filippo'*s 35th book, a story collection titled* The Great Jones Coop Ten Gigasoul Party, *will make its appearance this year from Wildside Press.*

natureOUTLOOK
TUBERCULOSIS

10 October 2013 / Vol 502 / Issue No 7470



Cover art: Neil Webb

Just when it seemed that humanity was ridding itself of its most lethal microbe, drug resistance and the HIV pandemic has kept *Mycobacterium tuberculosis* firmly on the map (S2).

To quell the rise of drug resistance, we need new types of drugs that act quickly and safely to increase the likelihood that a patient will finish their full course of treatment. In late 2012, the first new anti-TB drug in nearly half a century won market approval, and others should be soon to follow (S4).

Tuberculosis control also needs a practical point-of-care diagnostic – and one that also works in people with HIV. Our reporter visited clinics in Zambia to gauge the rapid roll out across Africa of the GeneXpert, and hears first hand the practical problems of using sophisticated technology in places with rudimentary facilities (S10).

These developments are the fruits of unprecedented public-private collaboration serving a market unable to afford market-priced medicines. To advance this enlightened effort, universities should learn to be less protective of intellectual property and promote affordable medicines for the world's poorest (S7).

While new technologies are important, we also need a broader concept of risk when considering disease susceptibility to help identify ways to alleviate the suffering inflicted by TB using existing technology (S13). Researchers also need to clarify how tuberculosis takes hold in a population and spreads (S16).

Achieving the ultimate goal of eliminating tuberculosis will require an effective vaccine. Ten years ago, no vaccine candidates were undergoing clinical testing; today there are more than a dozen (S8). While the recent failure of a promising vaccine in a clinical trial was a setback, researchers are taking heart from the fact it was possible to conduct the trial in the first place (pages S8).

We are pleased to acknowledge the financial support of Janssen Research and Development LLC in producing this Outlook. As always, *Nature* retains sole responsibility for all editorial content.

Tony Scully
Science Editor, Nature Outlook

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the Nature Outlook Editorial guidelines available at http://www.nature.com/advertising/resources/pdf/outlook_guidelines.pdf

CITING THE OUTLOOK
Cite as a supplement to *Nature*, for example, Nature **Vol XXX**, No. XXXX Suppl, Sxx–Sxx (2013). To cite previously published articles from the collection, please use the original citation, which can be found at the start of each article.

VISIT THE OUTLOOK ONLINE
The *Nature Outlook Tuberculosis* supplement can be found at http://www.nature.com/nature/outlook/tuberculosis

All featured articles will be freely available for 6 months.

SUBSCRIPTIONS AND CUSTOMER SERVICES
For UK/Europe (excluding Japan): Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242. Subscriptions and customer services for Americas – including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea:Nature Publishing Group — Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

CUSTOMER SERVICES
Feedback@nature.com

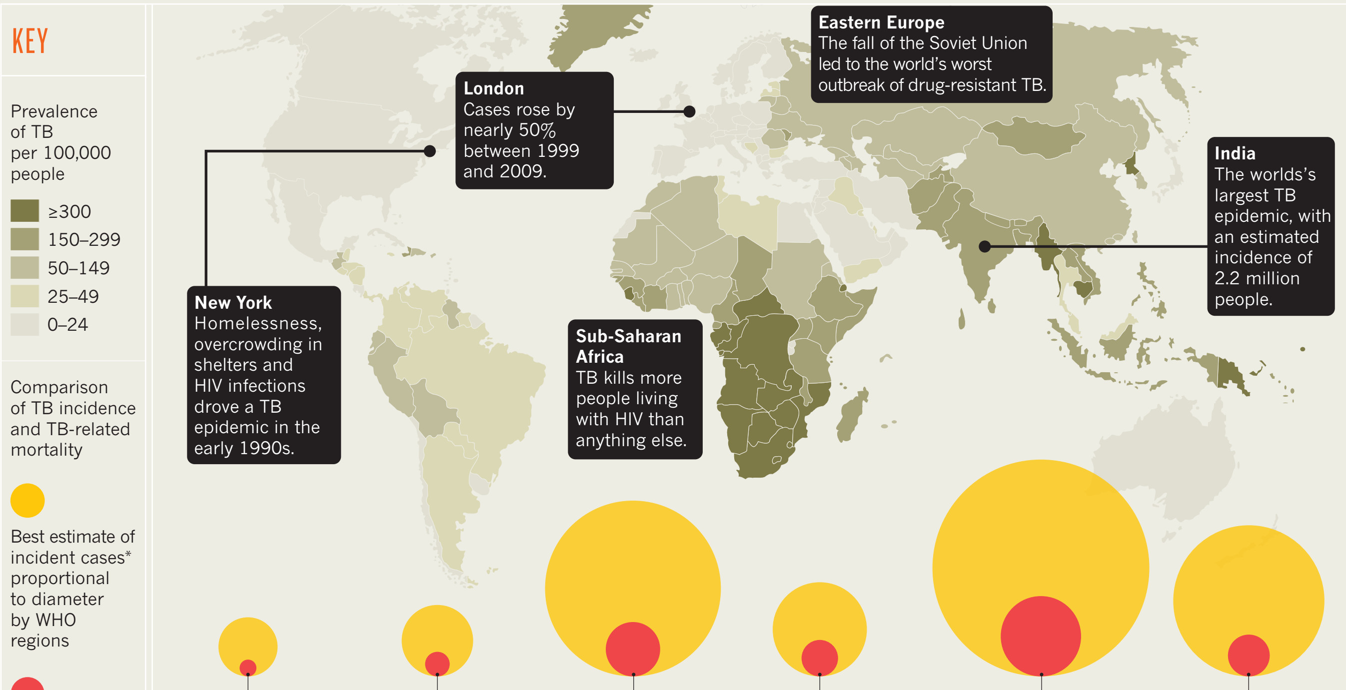
CONTENTS

A MORTAL FOE

Tuberculosis is one of the world's most lethal infectious diseases. Further progress in consigning it to the past is a massive challenge. By **Tom Paulson.**

GLOBAL BURDEN OF TUBERCULOSIS

In 2011, nearly 9 million people fell ill from TB and 1.4 million died, mostly in poor countries, with 60% of cases in Asia and 24% in Africa.



The Americas
Population: 943 M
Incidence: 260,000
Mortality: 21,000
HIV-positive (TB cases): 37,000

Europe
Population: 900 M
Incidence: 380,000
Mortality: 45,000
HIV positive: 23,000

Africa
Population: 857 M
Incidence: 2.3 M
Mortality: 220,000
HIV positive: 870

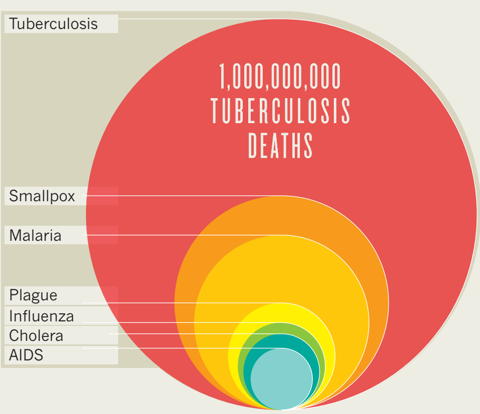
Eastern Mediterranean
Population: 609 M
Incidence: 660,000
Mortality: 99,000
HIV positive: 8,700

South East Asia
Population: 1. 8 B
Incidence: 3.5 M
Mortality: 480,000
HIV positive: 140,000

Western Pacific
Population: 1.8 B
Incidence: 1.7 M
Mortality: 130,000
HIV positive: 36,000

THE BIGGEST KILLER

Tuberculosis has killed more than any other infectious disease in history. Over a billion lives in the past two hundred years.

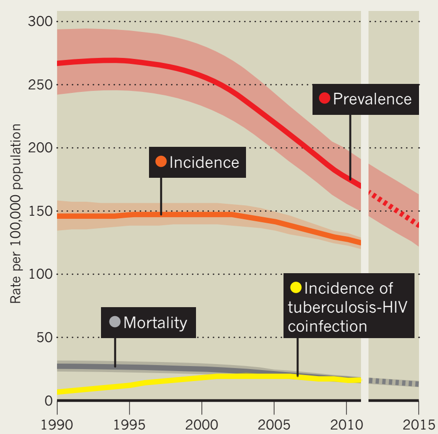


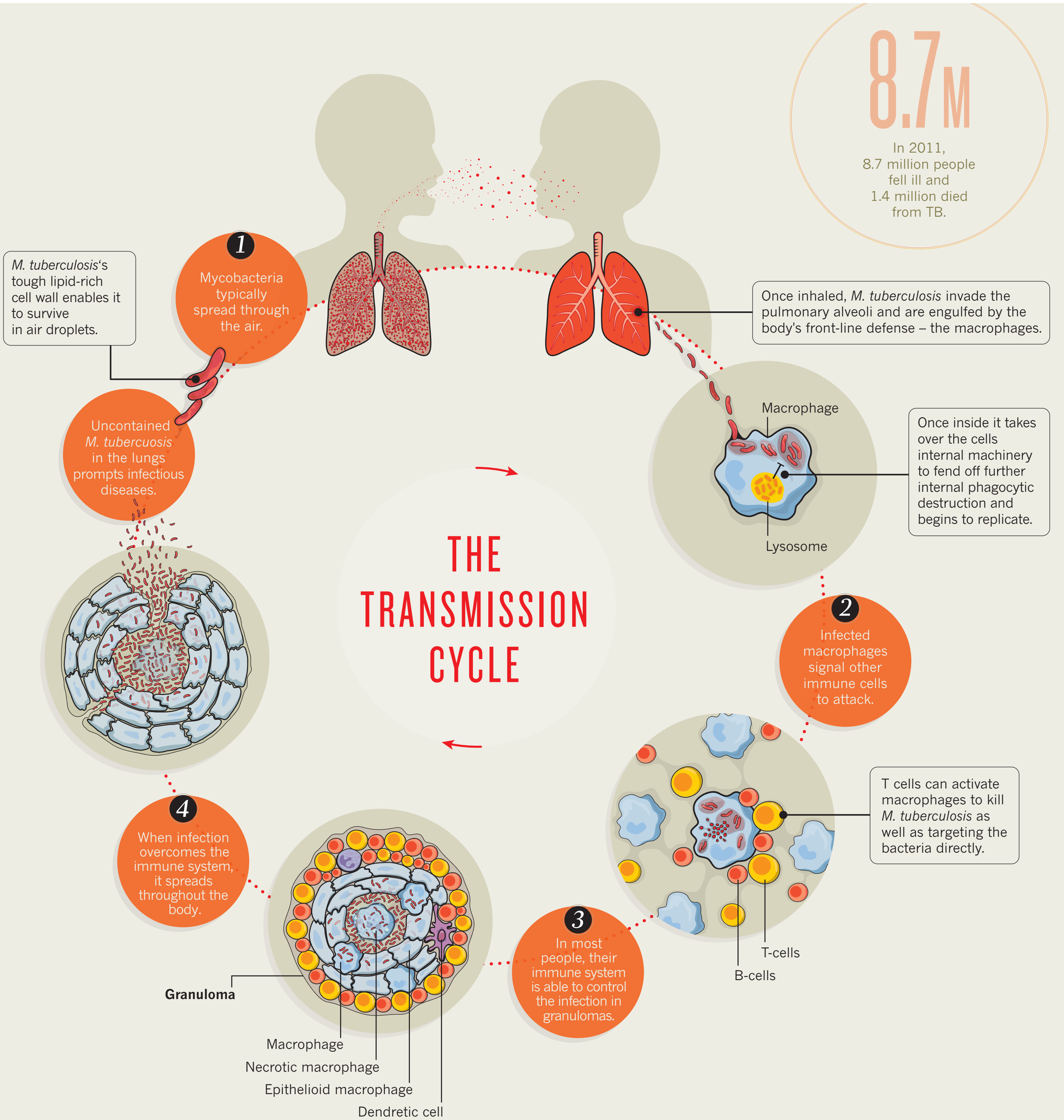
THE 100 YEARS BATTLE

Rising living standards in industrialized nations, interupted by two World wars, and new antibiotics had tuberculosis in decline.

SLOW PROGRESS

Over the past fifteen years, an invigorated anti-TB effort has begun to reduce the global burden of disease worsened by HIV.





1 Transmission

People with active disease develop symptoms such as a cough, which propels bacteria into the air where it can be inhaled by others. Tuberculosis must be diagnosed and treated as soon as possible to render the person non-infectious and prevent the spread of disease (see 'Control issues', page S16).

2 Immune response

The BCG vaccine offers little protection to adolescents and young adults from the form of the disease that causes most deaths. Any effective vaccine will need to harness T cells, but scientists are still looking for correlates of protection (see 'An age-old problem', page S8).

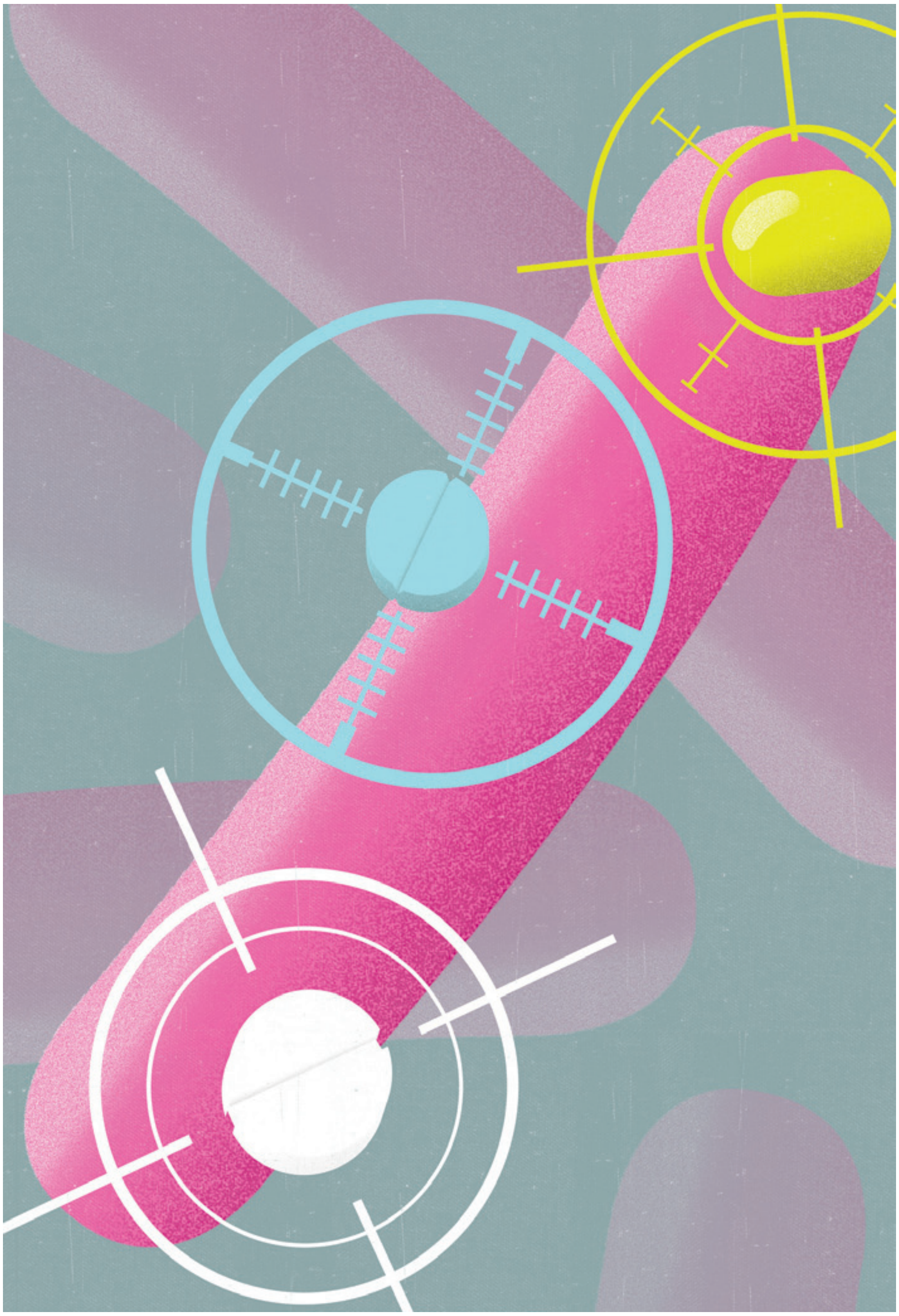
3 Latency

M. tuberculosis can be contained within granulomas for years. It is thought that latency may encompass a spectrum of states, from people who have completely controlled the disease, to those with undetected, subclinical disease (see 'Latency: A sleeping giant', page S14).

4 Activation

If the immune system weakens as a person ages or contracts HIV infection, for example, bacterial replication can overcome the immune system and granuloma breaks down, releasing *M. tuberculosis* into the lungs and advancing disease (see 'A combined effort', page S4).

WORLD TUBERCULOSIS REPORT 2012. LIENHARDT, C. ET AL. NATURE REVIEWS MICROBIOLOGY 10, 407-416 (2012)



NEIL WEBB

DRUG DEVELOPMENT

A combined effort

Combinations of anti-TB drugs are difficult to overcome because they attack Mycobacterium tuberculosis *in different ways.*

BY AMY MAXMEN

The ringing in Dalene von Delft's ears had become unbearable. The inescapable noise was caused by the drug amikacin, which von Delft — a doctor at Vergelegen Mediclinic hospital in Cape Town, South Africa — was taking to treat her drug-resistant tuberculosis (TB). But von Delft's biggest fear was that the ringing would go silent; in certain studies, up to half of those injected daily with amikacin for the two-year regimen were found to go deaf[1]. She wanted to cure her disease, but losing her hearing seemed like too high a price. As a doctor, "you need to be able to use a stethoscope", she says. "It would have ended my career." But stopping treatment altogether would have been an even more dangerous option. Without proper treatment, up to two-thirds of people ill with TB will die[2].

von Delft learned of a new anti-TB drug that was performing well in clinical trials. Although its safety had not been established, she pleaded with the drug's maker, Janssen Pharmaceuticals, based in Titusville, New Jersey, and the South African Medicines Control Council, and managed to switch to the unapproved drug, called bedaquiline. And although it made her heart palpitate temporarily, the remote danger of a heart attack seemed more bearable than losing her hearing. "You weigh the risks against the benefits," she admits. After two years, the TB was gone and her hearing remained. "I was very lucky," she says.

A new drug for TB has been a long time coming. Bedaquiline, approved by the US Food and Drug Administration (FDA) at the close of 2012, is the first novel anti-TB drug since rifampicin was introduced nearly half a century ago. With TB killing approximately 1.4 million people each year, industry, academic institutions and public-health organizations are having to collaborate like never before to stem the rise of drug-resistant TB.

NOT GOING QUIETLY

Rates of TB in the United States and Europe began to decline at the turn of the twentieth century because of sanatoriums built to isolate TB patients, along with improvements in overcrowded tenements that slowed TB transmission. Then, from the 1940s to the 1960s came antibiotics — streptomycin, isoniazid, rifampicin and 4-aminosalicylic acid (PAS) — that could effectively treat the disease. The arrival of these drugs led many to believe that TB would soon disappear altogether. "We were convinced in the 60s that we would end TB," recalls Jacques Grosset, a professor at the Center for Tuberculosis Research at the Johns Hopkins School of Medicine in Baltimore, Maryland, who is himself a victim of TB (he had a portion of his lung removed during a bout with the disease in the 1950s).

But those expectations failed to materialize. Many patients, lacking careful supervision, would stop midway through the one- to two-year-long course of treatment. When this

happened, or when the drugs were given singly rather than in combination, naturally drug-resistant strains of the TB pathogen thrived and overtook the more sensitive population. People feel better after a few months and stop taking their pills, says Grosset. Once a patient acquired drug-resistant disease, they could spread their resistant pathogens directly to their neighbours.

Then came the HIV epidemic, which made people acutely vulnerable to TB. Developing countries, which had never managed to fully suppress the disease, experienced skyrocketing TB rates, and Western countries saw a resurgence. In 1993, the World Health Organization (WHO) in Geneva, Switzerland placed the disease back on the public-health agenda by declaring TB a global emergency. New drugs were needed with different mechanisms of action, which would suppress resistance and shorten treatment times. And yet pharmaceutical companies had little incentive to invest the massive amounts of money necessary for drug development — TB predominantly affects people from low-income countries, who cannot afford expensive treatments. Without a business case for TB drug development, progress stalled.

In the wake of the WHO declaration, the Bill & Melinda Gates Foundation, headquartered in Seattle, Washington, along with other non-governmental organizations and government agencies, including the National Institutes of Health in Bethesda, Maryland, got together to convince drug developers across the public and private sectors to work together. In response, the pharmaceutical giants GlaxoSmithKline (headquartered in Brentford, UK), Novartis (Basel, Switzerland), AstraZeneca (London) and Sanofi (Paris) agreed to collaborate with each other and with universities to pick up where TB drug developers had left off in the 1960s.

SHAKY START

Developing a new drug for TB is a complex process. It involves screening thousands of compounds to find any that might kill the bacterium, *Mycobacterium tuberculosis*; adapting these compounds into substances that work as a drug; testing for efficacy in animal models; and finally testing how effective and safe the drug is in human patients.

Various labs were able to identify compounds that inhibited metabolic pathways and other vital processes occurring within the bacterium; unfortunately, it became apparent that many compounds selected through these screens had trouble penetrating the membrane of whole mycobacteria and subsequently failed tests in mice. So researchers changed their screening protocols to look for the effects compounds had on whole mycobacteria — even so, it has not been smooth sailing.

In 2009, a team at the Novartis Institute for Tropical Diseases in Singapore landed on a new class of compound, pyrimidine–imidazoles, which killed *M. tuberculosis in vitro* but had no effect on infected mice. In 2010, the researchers figured out why: the drug only blocked the ability of the mycobacteria to survive in its glycerol suspension — leaving it with little relevance in the world beyond the test tube[3]. "That was not fun," recalls Thomas Dick, who led the project at Novartis, and now directs the Antibacterial Drug Discovery Laboratory at the National University of Singapore after Novartis dropped out of TB drug discovery. "It was the failure of a two-year project that took a lot of investment."

"My fear is that what is really needed is orders of magnitude more funding."

Despite such disappointments, candidates for new anti-TB drugs are continuously being identified. In May 2013, for example, researchers reported that high doses of vitamin C wipe out cultures of *M. tuberculosis* by triggering a DNA-damaging reaction[4].

TB survivor van Delft took a chance on a new drug.

In July, a team at the Institute Pasteur Korea in Seongnam-si discovered a compound labelled Q203 that cuts off the energy supply to mycobacteria by blocking ATP synthesis both in culture and in mice[5].

Indeed, in the past decade, six types of compound that target *M. tuberculosis* in new ways have progressed to trials in humans. However, these trials face their own set of problems. Foremost among them is phase 2a testing, which checks whether a drug decreases the mycobacterial load in patients' sputum within two weeks of starting treatment. Because some drugs act slowly, the short timeframe of this trial can be misleading. In fact, a number of people at Janssen wanted to shut down the bedaquiline programme after it performed poorly in a phase 2a trial, says Myriam Haxaire-Theeuwes, who is developing bedaquiline at Janssen's research and development branch in Beerse, Belgium. Spectacular results in cell culture and in mice convinced the team of bedaquiline's worth. "You need to have strong product champions," she says, "and senior management that will listen."

NATURE.COM
For a review of tuberculosis drug discovery, visit:
go.nature.com/ruw6so

Phase 2b trials are problematic for a different reason. For drug-resistant TB, such a trial needs to run for at least two years to determine whether the new treatment works: it can take that long to cure the disease. In the meantime, at least 310,000 patients with drug-resistant TB are taking potentially toxic medicines each year[2].

To address the urgent need for new treatments for TB, the FDA announced in 2009 that in some cases they would grant 'accelerated approval' for promising drugs. For example, treatment success could be measured by no detectable mycobacteria in a patient's sputum after six months of treatment, rather than ensuring that patients fully recover from drug-resistant TB at the end of a 2.5 year study.

It was under these new guidelines that bedaquiline was approved for multidrug-resistant TB in late 2012. Shorter clinical trials, however, leave more unknowns, including rates of cure and the incidence of side effects. "In situations where patients have few treatment options, healthcare providers will accept greater risk," explains Edward Cox, the director of the FDA's Office of Antimicrobial Products.

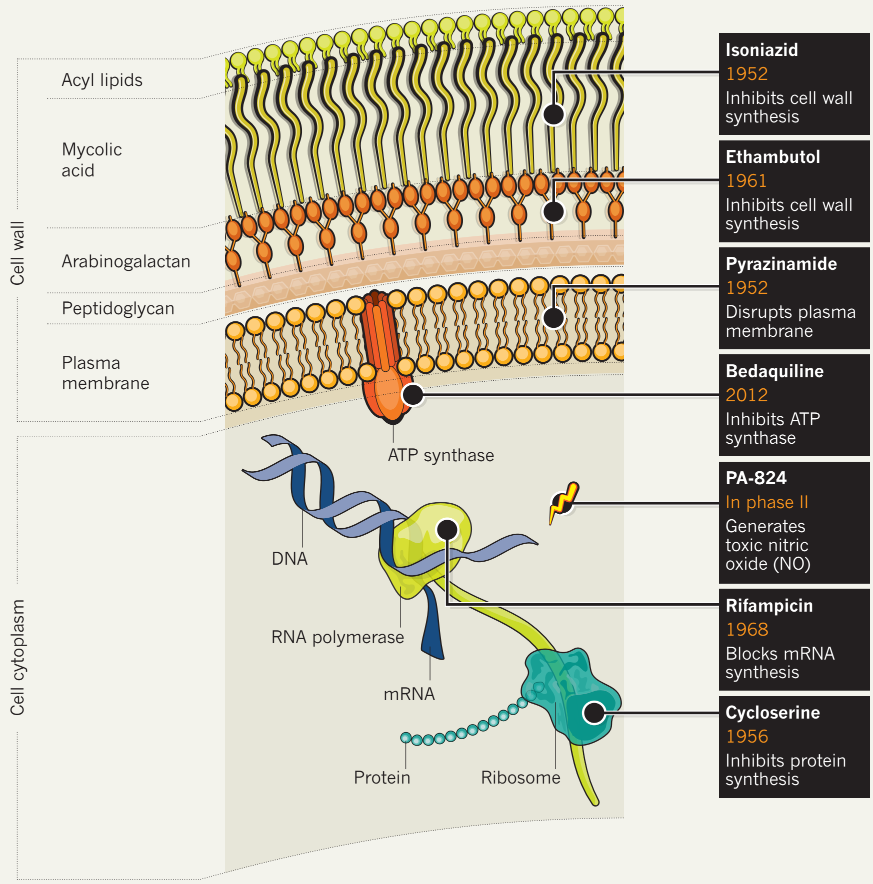
In June 2013, Janssen began to sell bedaquiline in the United States, and approval is pending in several other countries. In the meantime, the drug is undergoing a phase 3 trial to assess with greater certainty its effects on drug-resistant TB. The Global Alliance for TB Drug Development (the TB Alliance), a non-profit organization based in New York, is also running a trial on bedaquiline — this time in combination with a leprosy drug, clofazimine, and another novel anti-TB agent, PA-824, in hopes of reducing the duration of the current six-month course of treatment for drug-susceptible TB.

JOINING FORCES

Although they are often used as a marker of how well a drug is doing in trials, sputum bacterial counts are far from ideal. Counts can vary from day to day in a single patient, making the measure rather messy, says Clifton Barry, a TB researcher at the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, and at the International Tuberculosis Research Center in Masan, South Korea. In his Korean laboratory, Barry scans patients' lungs with computed tomography (CT) and positron emission tomography (PET) to see

HIT LIST

Anti-TB drugs attack *Mycobacterium tuberculosis* in different ways, so combining different drugs helps combat the bacterium developing ways to overcome a drug.



how drugs alter the state of *M. tuberculosis* inside the patient. Because the bacteria lurk in patients' lungs long after they leave their sputum, he says this method is far more precise.

Scans may also help reveal complementary drug combinations. Thomas Dick in Singapore says that the ideal drug combo would include medicines that quickly kill *M. tuberculosis* replicating in the lung fluid, along with slower acting drugs that hit hard-to-reach mycobacteria hiding out in lesions in the lungs.

To find drugs that penetrate lesions, Dick proposes that researchers switch some of their animal studies to rabbits because, unlike in mice, *M. tuberculosis* forms lesions in their lungs. "It allows for a more rational selection of compounds that will work well in combination," Dick says.

One new drug with a novel mechanism is probably not enough to fight *M. tuberculosis* drug resistance. To tackle this, the Bill & Melinda Gates Foundation funds the TB Drug Accelerator programme, which partners seven drug companies with several laboratories at publically funded institutions. "We want at least one combination of three agents that every TB patient is sensitive to so that the current notion of drug resistance just goes away," says Ken Duncan, deputy director of drug discovery at the Gates Foundation.

The companies involved in the programme share their compound libraries and their results. Barry, who is a participant, says that sharing tools and information at each step of drug development eliminates the redundancy that can occur in more secretive and independent drug development programmes where higher profits are at stake.

The FDA is also looking at ways to speed up the development of combination therapies. In 2010, they issued draft guidelines for testing combinatorial treatments in urgent situations. These guidelines will permit researchers to skimp on data about the effects of individual components within new combinations in cases where the drugs are for the treatment of life-threatening diseases for which there are no satisfactory alternatives.

"You need to have strong product champions."

The coming decade looks bright for TB drug development. By 2014, the European Medicines Agency in London is expected to approve delaminid, a novel drug from the Japanese drug company Otsuka, based in Tokyo. This compound, a nitroimidazole, poisons mycobacteria by releasing nitric oxide once it is metabolized. In July 2013, Pfizer, headquartered in New York, sold a novel drug candidate — sutezolid — that was sitting on its shelf to Sequella, a small pharmaceutical company in Rockville, Maryland, for development and commercialization. Sutezolid, a drug that prevents mycobacteria from making proteins, looked promising in mouse studies, but Pfizer had frozen its development at that stage[6].

MONEY MATTERS

Early results from the TB Alliance trial on bedaquiline, clofazimine, PA-824 and pyrazinamide also look encouraging. Trials like this are exceptional in the drug development world because they involve compounds owned by multiple companies. Khisimuzi Mduli, a drug development project leader at the TB Alliance, suggests that the strength of the TB Alliance is that they have nothing to gain financially from drug development. This means that "our partners in pharma allow us to use their compounds in long clinical trials", he says.

Even as such collaborations bring hope to the TB research community, uncertainty looms. As the founding director of the new KwaZulu-Natal Research Institute for Tuberculosis and HIV, Durban, South Africa, William Bishai is both thrilled with the recent infusion of money for TB research and filled with dread that it is not enough. "My fear is that what is really needed is orders of magnitude more funding rather than slight increases of 10 to 20 per cent per year," says Bishai, who recently stepped down from the position to return to Johns Hopkins in Baltimore, Maryland.

Working in South Africa put Bishai at the heart of the TB epidemic. As well as giving his team access to a large pool of patients, he also made a happy discovery: there is no shortage of people willing to go after the disease. Like Dalene von Delft and Jacques Grosset, their personal experience of TB is motivation enough. "When I give a lecture here, two-thirds of the room has had TB or knows someone who has died from TB," he says. "So there's a real fire in the belly of the young people here, and they want to be part of the next generation of scientists fighting TB." ■

Amy Maxmen *is a science writer based in Brooklyn, New York.*

1. Frymark, T. *et al. Evidence-Based Systematic Review: Drug-Induced Hearing Loss—Amikacin* (American Speech-Language-Hearing Association, 2010).
2. World Health Organization. *Tuberculosis Factsheet No. 104* (World Health Organization, 2013).
3. Pethe, K. *et al. Nature Commun.* **1,** 57 (2010).
4. Vilchèze, C. *et al. Nature Commun.* **4,** 1881 (2013).
5. Pethe, K. *et al. Nature Med.* **19**, 1157–1160 (2013).
6. Lessem, E. *The 2013 Pipeline Report* 22–261 (i-Base/TAG, 2013).

PERSPECTIVE


Graduation time

Universities should forego profits from tuberculosis, say **David G. Russell** and **Carl F. Nathan**.

Tuberculosis (TB) is the single leading cause of death from bacterial infection. It is rapidly becoming untreatable, and untreated TB has a fatality rate of about 70% after three years. The challenges in developing new drugs for TB are scientific, logistical, fiscal and societal.

Over the past decade many pharmaceutical firms have abandoned antibiotic research, having failed to discover effective candidates with new mechanisms of action. A further disincentive is the lower return on investment that rapidly curative drugs offer compared with palliative medications for prevalent conditions. The financial picture is particularly bleak for TB, which chiefly afflicts people in low- and middle-income countries. The prospect of even scantier profits makes it all the harder to entice drug companies to work on TB rather than infections that are common in wealthy markets.

The treatment of TB requires combination chemotherapy, because the use of a single agent virtually guarantees the rapid emergence of resistance. When new TB drugs do reach the market, rampant drug-cutting and counterfeiting in poorer countries mean that one or more of the drugs in the combination may be absent, or present at suboptimal concentrations, promoting the accelerated emergence of drug resistance, which is already prevalent to each of the widely used TB drugs. Consequently, one new TB drug is unlikely to do the trick: we need sets of drugs that work together. The difficulty of finding a new combination is more than additive, as each drug must not interfere with the others or with the antiretrovirals used to treat HIV infection (a common co-infection in sub-Saharan Africa). Thus, even if a pharmaceutical company discovers one effective new TB drug (a big if), the chances are that such a drug would be rapidly lost to resistance unless it were used in combination with two or three other new drugs. Embarking on such a search takes an unprecedented level of social consciousness and cross-industry cooperation.

THE FINANCIAL PICTURE IS PARTICULARLY BLEAK FOR TB.

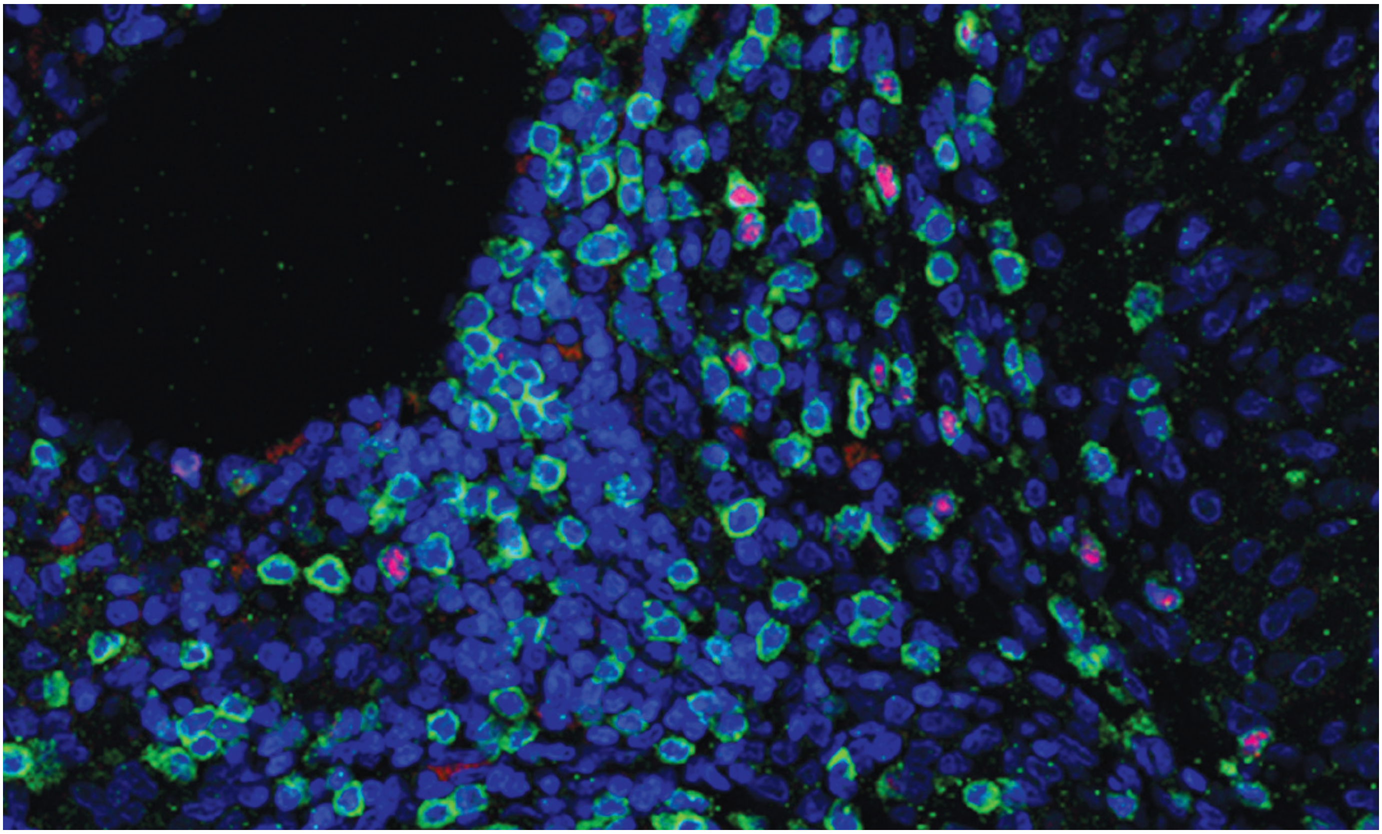
Despite these formidable challenges, a remarkable number of public–private partnerships (PPPs) have been launched for TB drug discovery and development. Among them are the Global Alliance for TB Drug Development, headquartered in New York, the Bill & Melinda Gates Foundation's TB Drug Accelerator (Seattle, Washington), the Lilly TB Drug Discovery Initiative (Seattle), the Tres Cantos Open Lab Foundation (Guildford, UK), and the Innovative Medicines Initiative (Brussels). Many of these consortia pair academic researchers who have an up-to-date understanding of *Mycobacterium tuberculosis* biology and TB pathogenesis with pharmaceutical scientists who have access to chemical compound collections that are larger than those in universities, more suited to drug development, and more extensively curated. Even more important, the pharmaceutical companies bring expertise in medicinal chemistry, chemoinformatics, pharmacology, pharmacokinetics and toxicology, along with the infrastructure to perform clinical trials. Such partnerships provide outstanding opportunities for innovation and efficiency.

Although these PPPs commonly have more open approaches to intellectual property (IP) than are usual in drug development, several pharmaceutical companies have proved willing participants in the search for treatments for neglected diseases. Unfortunately, this enlightened attempt to find a solution to the depleted TB drug pipeline has not been matched by all universities. Many academic institutions are struggling to fill a fiscal deficit. In the United States, as a consequence of the Bayh–Dole Act of 1980, universities in receipt of federal funding are free to license IP as a source of income. Under this profit-driven model, some institutions prefer to save precious patent-filing funds for IP with greater potential return. Some do not want to commit to the distribution of drugs on a non-profit basis in the public markets of low-income countries, a condition required by many of the PPPs. There may also be concerns about sharing ownership of IP with a drug company, or universities may demand a share that is disproportionate with their contribution. Few institutions have heeded the call of the student Universities Allied for Essential Medicines, a global group headquartered in Oakland, California, that challenges universities to adopt IP policies that promote affordable access to medicines and medical technologies for the world's poor.

Although the technology transfer offices in universities may regard the potential income from the IP associated with the early stages of drug discovery as an attractive source of funds, this expectation is unrealistic for TB drug development because of the limited commercial return. Moreover, any such income pales in comparison with the hundreds of millions of dollars that need to be invested to turn lead compounds into drugs. Who will pay to develop the best candidates that emerge from the PPPs? To conduct the clinical trials? To deliver the drugs to those in need? Regrettably, both a penchant for fiscal conservatism and a zeal for IP protectionism at some universities can obstruct the earliest stages of drug discovery.

The world cannot afford to wait long for answers to these questions, and universities need to play their part in finding solutions. For example, they will need to absorb some of the costs of IP protection and accept the potential for low fiscal returns for IP related to drug discovery for diseases like TB. More broadly, universities should lobby for larger, lasting fixes to the broken antibiotic pipeline, such as those called for by a recent panel of the Institute of Medicine of the US National Academies, based in Washington, DC. We need to examine ways of uncoupling the direct link between the rewards for antibiotic development and income from selling the drugs. As an alternative solution, a global fund — perhaps capitalized by financial transaction taxes — could compensate participating drug discoverers for their activities in proportion to their products' reduction of the burden of disease. This would encourage industry and academia to become partners, promote cooperation, and render counterfeiting non-profitable. At this critical time it is imperative that universities re-evaluate their position and become activists for global health. ■

David G. Russell *is the William Kaplan Professor of Infection Biology in the Department of Microbiology and Immunology at the College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA.* **Carl F. Nathan** *is the R.A. Rees Pritchett Professor of Microbiology and Chair of the Department of Microbiology and Immunology at the Weill Cornell Medical College, New York, New York 10065, USA.*



JOSHUA MATTILLA/UNIVERSITY OF PITTSBURGH

T cells (green) and nuclei (blue) in a tuberculous granuloma stained for regulatory T cells are important regulators of macrophage activation.

VACCINES

An age-old problem

Researchers are on the hunt for a better alternative to the BCG vaccine.

BY SARAH DEWEERDT

It is one of the most widely administered vaccines in the world. In use for nearly a century, it has been given to over a billion people. And yet, it is still not good enough.

The bacillus Calmette–Guérin vaccine, or BCG, is a live but weakened bacterium that is about 80% effective at protecting young healthy children from severe forms of tuberculosis (TB) — particularly tuberculous meningitis, which affects the brain, and miliary TB, a systemic infection of the body.

But BCG offers little protection to adolescents and young adults from the form of the disease that spreads widely and causes most deaths: pulmonary TB. Moreover, BCG is a replicating bacterium; in children infected with HIV the vaccine itself can cause disease, known as BCG-osis.

Many experts say that eradicating TB will not be possible without an improved vaccine that protects against pulmonary TB. Because of drug resistance and the complexity of TB treatment regimens, "It is very clear that by far the most sustainable intervention would be if you have a vaccine that prevents disease," says Willem Hanekom, director of the South African Tuberculosis Vaccine Initiative (SATVI) in Cape Town.

At first glance, the effort to develop an effective vaccine against pulmonary TB seems poised to be richly rewarded. More than US$95 million went towards TB vaccine development in 2011 alone, and over a dozen candidates are in clinical trials. Just over ten years ago there were none.

The vaccine candidates are not only numerous but varied, using different mechanisms and dosing strategies. But this multitude of possibilities also reflects the field's biggest problem: a lack of knowledge about *Mycobacterium tuberculosis* and how it interacts with the human immune system. "We don't know enough to exclude anything at the moment," says Helen McShane, a vaccinologist at the University of Oxford, UK.

These challenges were highlighted by the failure of one promising vaccine candidate, MVA85A, to show efficacy against TB in a clinical trial in early 2013.

MVA85A is designed to stimulate immunity to 85A, one of several thousand molecules, or antigens, produced by *M. tuberculosis*. The vaccine delivers the antigen into cells — where it can trigger an immune response — via a genetically engineered virus, so MVA85A is known as a viral vector vaccine.

The research team conducting the trial, including Hanekom and McShane, gave MVA85A injections to 1,399 South African infants who had been vaccinated with BCG earlier in infancy[1]. Several vaccine candidates are designed to work this way, known as a prime–boost strategy. It is thought that priming the immune system with BCG and then administering a booster dose of a new vaccine will continue to protect children against the most dangerous forms of TB while adding protection against pulmonary TB.

But, after about two years, children who had received the MVA85A vaccine had the same rate of TB infection and disease as children who had not[1]. "That it showed no efficacy is surprising, because this vaccine was vetted in many ways," says Daniel Zak, a principal scientist at the Seattle Biomedical Research Institute in Washington.

Mouse and non-human primate studies had shown that MVA85A provides increased protection against TB over BCG alone; it is not clear

why these results did not translate to humans.

"Tuberculosis is a complex pathogen," says McShane. "It hides inside cells most of the time, which means the immune response required to clear it is predominantly a T-cell response." T cells are involved in a process called cellular immunity, in which, among other things, they induce infected cells to 'commit suicide'. By contrast, most of the successful vaccines in use today are based on humoral immunity, which mainly involves the production and secretion of antibodies into the extracellular fluid. Scientists have a better understanding of humoral immunity and the vaccines that stimulate it, but on its own it does not clear bacteria from cells.

TRIED AND TESTED

Despite the lack of efficacy shown in the MVA85A trial, researchers say it was a success in other ways. It was the first efficacy trial of a TB vaccine in more than four decades. It demonstrated that a TB vaccine can be tested in infants according to modern-day ethical standards, and that researchers can develop the infrastructure and analytical methods needed to carry out a large clinical trial in a resource-poor country[2]. "In a way it's become a template for how to do such a study," says Hanekom.

With so many vaccine candidates waiting in the wings, it is a template that could get a lot of use. The next vaccine candidate likely to start efficacy trials is known as M72. It is an example of what scientists call a subunit or protein-adjuvant vaccine: a cocktail of two antigens, 32A and 39A, plus an adjuvant, a chemical that enhances immune responses. As with MVA85A, scientists envision using M72 as a booster vaccine in individuals who have already received BCG.

Unlike the MVA85A trial, which was confined to infants, a team including SATVI researchers plan to test M72 efficacy in young adults, because they are more affected by pulmonary TB. So far, the team has shown that the vaccine is safe and induces a good T-cell response when administered to adults in South Africa either without TB or with latent infection[3]. An upcoming study will test whether the vaccine reduces the risk of developing active disease over the course of three years. It will involve around 7,000 participants in three African countries, Hanekom says, and is likely to begin by the end of 2013.

This is the first protein-adjuvant vaccine candidate to be tested for efficacy, notes Jelle Thole, executive director of the Tuberculosis Vaccine Initiative in Lelystad, the Netherlands, which supports and coordinates TB vaccine research throughout Europe. It was chosen largely because it happens to be the vaccine in this category that was the most developed. But in the future, Thole adds, vaccine candidates within each category should be tested against each other, with the one that does best in animal studies and at inducing immune responses in healthy volunteers moving forward first to efficacy trials.

A vaccine of a third type, based on a whole, live bacterium, may also be nearing efficacy trials. "This is one of the few candidates that is trying to be better than BCG," says Stefan Kaufmann, director of the department of immunology at the Max Planck Institute for Infection Biology in Berlin.

Kaufmann is one of the developers of the vaccine, called VPM 1002. It is based on BCG, which in turn is derived from *M. bovis*, which causes TB in cows. He notes that other types of vaccine contain only one or a few antigens to stimulate an immune response. His team chose to work with a whole live bacterium because "we thought that the whole antigen repertoire that is present in BCG might be a better choice," he says.

If it works, Kaufmann hopes that VPM 1002 will one day replace BCG as the standard immunization given to infants. It could also be combined with one of the booster vaccines now in development.

To create VPM 1002, Kaufman's team inserted a gene called listeriolysin into BCG and deleted another gene, encoding a subunit of urease[4]. These genetic changes make BCG more visible to the immune system. For example, BCG mainly stimulates CD4 T cells whereas VPM 1002 is better at also stimulating CD8 T cells. Kaufmann argues that a broader immune response is more likely to result in improved protection from the different forms of the disease.

"We don't know enough to exclude anything at the moment."

VPM 1002 has bested BCG in mouse studies, and its safety has been assessed in newborns in South Africa. The results of that study have not yet been published, but "everything went very well," Kaufmann says. The next step is to secure approval — and funding — to test the vaccine's efficacy.

GOING ROGUE

Meanwhile, some scientists are beginning to rethink what a successful TB vaccine might look like. For example, the Seattle-based Bill & Melinda Gates Foundation, which has contributed hundreds of millions of dollars to TB vaccine research, has issued a call to develop vaccines that rely on mechanisms other than T cells, recruiting parts of the immune system such as natural killer cells or antibodies that are not big players in the body's normal immune response to TB. That is likely to be a difficult exercise, but Zak thinks it might be worth trying, as "there's something that's missing" from the immune response induced by T-cell vaccines so far.

McShane, on the other hand, is not giving up on T-cell-inducing vaccines, or indeed on MVA85A. She is studying whether MVA85A will work if delivered directly to the airways as an inhaled vaccine. "We need to generate more potent vaccines. Delivering vaccines to the lung appears to be one way to do that," she says. Her lab is also investigating whether combining MVA85A with another molecule designed to boost the immune response will improve efficacy.

Others are reconsidering not just how but when the vaccine should be delivered. For example, some investigators are asking whether a booster dose of BCG itself could improve protection against pulmonary TB. And an effective vaccine may require even more frequent dosing. "I can imagine a scenario where you need to take a puff of your TB vaccine on your birthday," says Kevin Urdahl, a principal investigator at the Seattle Biomedical Research Institute.

How to prioritize clinical trials and translate basic research more effectively into vaccine development is also generating interest.

"The big issue in vaccination right now is that we don't have a good correlate of protection," says Thole. In other words, there is no test or measurement that can define whether or not a person is protected from TB. "So we don't know what kind of immunity a vaccine needs to induce."

Researchers know that an effective TB vaccine must induce TB-specific T cells, but this does not seem to be the whole story. Take the MVA85A trial, in which researchers observed a T-cell response to the vaccine despite its lack of efficacy. Ideally, additional correlates of protection would be identified by comparing immune responses in individuals who are protected by a vaccine to those in individuals who are not protected.

The negative results of the MVA85A trial mean that researchers cannot make this comparison, so this strategy remains elusive. Still, McShane hopes to mine the trial data to see if there are any differences in the immune responses between study participants who became infected with TB and those who did not.

Others are looking not for correlates of protection but for what they call correlates of risk. By following the natural history of TB from the time of infection, "we can comprehensively define the immune responses in people that progress to [active] disease and those who do not progress", says Zak, who is conducting such studies in collaboration with SATVI.

Researchers could then conduct clinical trials of new vaccine candidates in people at high risk of progressing to active disease (see 'Latency: A sleeping giant' page S14), boosting the statistical power of smaller, more rapid studies.

The need for a better clinical trial strategy is acute, some researchers say. "My fear is that if we push the clinical trials too quickly, and spend a lot more money getting results [from vaccines] that don't provide much protection, then we'll have lost our chance and the resources will run out," says Urdahl. "There's more basic research to do on the front end to make sure that we're putting the right candidates into the pipeline." ■

Sarah DeWeerdt *is a science writer based in Seattle, Washington.*

1. Tameris, M. D. *et al. Lancet* **381,** 1021–1028 (2013).
2. Tameris, M. *et al. Tuberculosis* **93,** 143–149 (2013).
3. Day, C. L. *et al. Am. J. Respir. Crit. Care Med.* **188,** 492–502 (2013).
4. Ottenhoff, T. H. & Kaufmann, S. H. *PLoS Pathog.* **8,** e1002607 (2012).

CATHERINE DE LANGE

The health clinic in Kanyama, a slum district of Lusaka, Zambia, is trying out a new device to test for tuberculosis.

DIAGNOSIS

Waiting for results

There are several new tests for tuberculosis in the pipeline, but they must be shown to be effective in areas with limited resources and a heavy burden of HIV.

BY CATHERINE DE LANGE

Tobias Hamooya says he is feeling a bit better these days, but even to the casual observer it is clear that he is still not doing well. Sitting on the porch outside the Macha Mission Hospital in rural Zambia, his slow, staggered speech is punctuated by a violent cough, and just talking seems to leave him drained.

A few weeks ago Hamooya's cough got so bad he left his wife to look after their newborn baby and six other children and travelled the 150 kilometres to get here. Did he have any idea what was wrong with him? His answer needs no translation: "TB". Since arriving at the clinic they also tested him for HIV, and the results came back positive for that too. Once his tuberculosis (TB) medication kicks in, he will begin taking antiretrovirals (ARVs).

Pretty much everyone in the region is infected with TB, says John Spurrier, medical adviser at the hospital where Hamooya is being treated. "In most people your body handles it. You may go all your life and never have a problem. Now with HIV, which destroys the immune system, it becomes a big issue."

The highest proportion of new TB cases is in sub-Saharan Africa: more than 260 people per 100,000 in 2011. By comparison, in the same year France saw 4 cases per 100,000. And the region is in the grips of an HIV epidemic; TB kills more people living with HIV than anything else, and detection and treatment of TB is vastly complicated by HIV co-infection. Where once treatment programmes operated independently, many countries like Zambia now try to test and treat the two together. In 2004, just 3% of TB patients in the World Health Organization (WHO)'s African Region were tested for HIV; in 2011, it was 69%. Hamooya was one of the lucky ones: detecting TB is extremely difficult in patients who also have an HIV infection.

HIV has an effective point-of-care test that can quickly and accurately detect infection and is practical and cheap enough for use in rural clinics. TB does not. In recent years, finding such a test for TB has been a major public-health goal, and these efforts are beginning to bear fruit. Over the next few months, a new test called Xpert MTB/RIF that detects TB DNA will enter clinics in several high-burden countries in southern Africa; other tests are close behind. The question is, in the uncompromising settings that are home to the greatest burden of disease, will these new tests fulfil their potential?

NEED FOR SPEED

A fast TB diagnosis is important. The sooner a patient is diagnosed, the sooner they can start treatment to mitigate the debilitating symptoms of TB and limit the potential for transmission. But there is an added imperative for HIV-positive patients. "Starting someone on ARVs who has [untreated] TB means that, as their immune system recovers, they can get very sick," says Spurrier.

Spurrier is referring to IRIS — immune reconstitution inflammatory

➲ NATURE.COM
For more on the simple fluorescent dye test for TB, visit go.nature.com/wmtneq

syndrome. "Most of the symptoms you have from disease are [from] your body fighting the disease, not from the virus itself. So if your immune system is not working, you can have disease all through your body and have relatively few symptoms," he says. Once the rebuilt immune system tries to fight TB, the immune response can be lethal.

The standard method for testing for TB in the clinic, smear microscopy, is not very fast. The technique has changed little since it was developed by German scientist Robert Koch in the 1880s; it involves staining and examining a sample of sputum — the mucus patients cough up — under a microscope, counting the bacteria and grading the severity of the infection. Although simple, this technique can take as long as a week to get a result and heavily depends on the experience of the technician performing the test. But even then it might be inconclusive: the test is estimated to detect TB in only around 40–50% of people with HIV[1], rendering it impractical for parts of the world without the capacity for further tests. "Sputum is not very accurate: we miss people and then need to do additional work using X-rays and so on," says Nzali Kancheya, TB programme director at the Center for Infectious Disease Research in Zambia (CIDRZ) in Lusaka.


CATHERINE DE LANGE

Tuberculosis patient Tobias Hamooya.

HIGH HOPES

Xpert MTB/RIF aims to address these shortfalls, and its roll out is causing a big buzz. It is a fully automated machine and cartridge system that uses polymer chain reaction (PCR) to amplify and detect DNA sequences specific to *Mycobacterium tuberculosis* in sputum. The technology uses the GeneXpert system, an existing molecular testing device developed by molecular diagnostics company Cepheid, based in Sunnyvale, California. Cepheid, together with the Foundation for Innovative New Diagnostics in Geneva, Switzerland, and researchers from the University of Medicine and Dentistry of New Jersey, with additional funding from the National Institutes of Health in Bethesda, Maryland, developed the TB-specific system in 2008.

"It's made to be used by people with minimal training, even though it's a high-tech piece of equipment," says Monde Muyoyeta, one of the lead researchers on the Zambia AIDS Related Tuberculosis (ZAMBART) Project, a non-governmental organization formed through a collaboration between the University of Zambia's School of Medicine in Lusaka and the London School of Hygiene and Tropical Medicine. Sputum samples are loaded onto cartridges and inserted into the machine, which is connected to a computer to read off the results. In theory, this set-up should make it easy to diagnose patients, as the machine does all the analysis.

In December 2010, WHO recommended the use of Xpert MTB/RIF to diagnose TB and is monitoring a massive global roll out. Part of this comes in the form of the TB Xpert Project, a US$25.9 million programme funded by UNITAID, which is hosted by WHO in Geneva, and will provide over 220 machines and 1.4 million cartridges to 21 countries, mainly in East Africa and South-East Asia. Many countries are able to buy the system at a discount; as of March 2013, more than 6,100 machines and 2.3 million cartridges had been purchased[2] (see 'Xpert sales', page S12).

In Zambia, several Xpert machines have been bought for use in clinics, in anticipation of government guidelines due later this year. In the meantime, other machines are already in use for research, for instance by ZAMBART, to look at how effective it will be in resource-poor settings. A lot rests on this technology, says Kancheya. "In Western countries they have a lot of other technologies, but in places like Zambia, GeneXpert will really help us," she says. But first, solid evidence about its utility needs to be collected from the field, with the ZAMBART Project taking the lead.

> ***"Sputum is not very accurate: we miss people and then need to do additional work."***

PRACTICAL EXAM

The health clinic in Kanyama is a bright spot in this densely populated slum in the west of Zambia's capital, Lusaka. It is also one of two sites chosen by ZAMBART to test GeneXpert in a primary healthcare setting. In addition to the GeneXpert units, the clinic is home to a make-shift chest X-ray clinic, housed away from the dust in a shipping container painted with scenes of TB care in Africa. Kanyama sees the highest rates of TB in the city, and the clinic is the ideal arena to study how the GeneXpert technology performs far from the sterile lab.

For example, GeneXpert depends on a continuous supply of electricity. "If you don't have that, how does it affect the performance of the tests?" says Muyoyeta. Regular power cuts are a reality in Kanyama. "With smear microscopy, if you run out of electricity you can just come back and finish your test later," she says. "But with GeneXpert, if you have a power outage and you don't have backup, you have to start all over again." If you are lucky, you have some leftover sample; if not, you will have to call the patient back to provide another. These issues are crucial in determining whether the system can perform as hoped. "If you're wasting samples and not getting the result the same day then it's not a point-of-care test," says Barry Kosloff, lab manager on the ZAMBART Project.

Then there is the cost. Under concessional pricing, the GeneXpert machine plus a computer to display the results costs around US$17,000. The cartridges cost just under $10 each. You need one cartridge per test, and the simplest version of the machine can process four cartridges at a time.

The cost is significant, Muyoyeta says. By comparison, "a microscopy slide is less than one dollar." And, when you add in the capital cost and additional costs like electricity, GeneXpert will be far more expensive. Technology like this might leave low- or middle-income countries more dependent on outside help. "I think it can make a difference," says Muyoyeta, "but it's so expensive there's no way that the minister of health can afford to roll out GeneXpert without support."

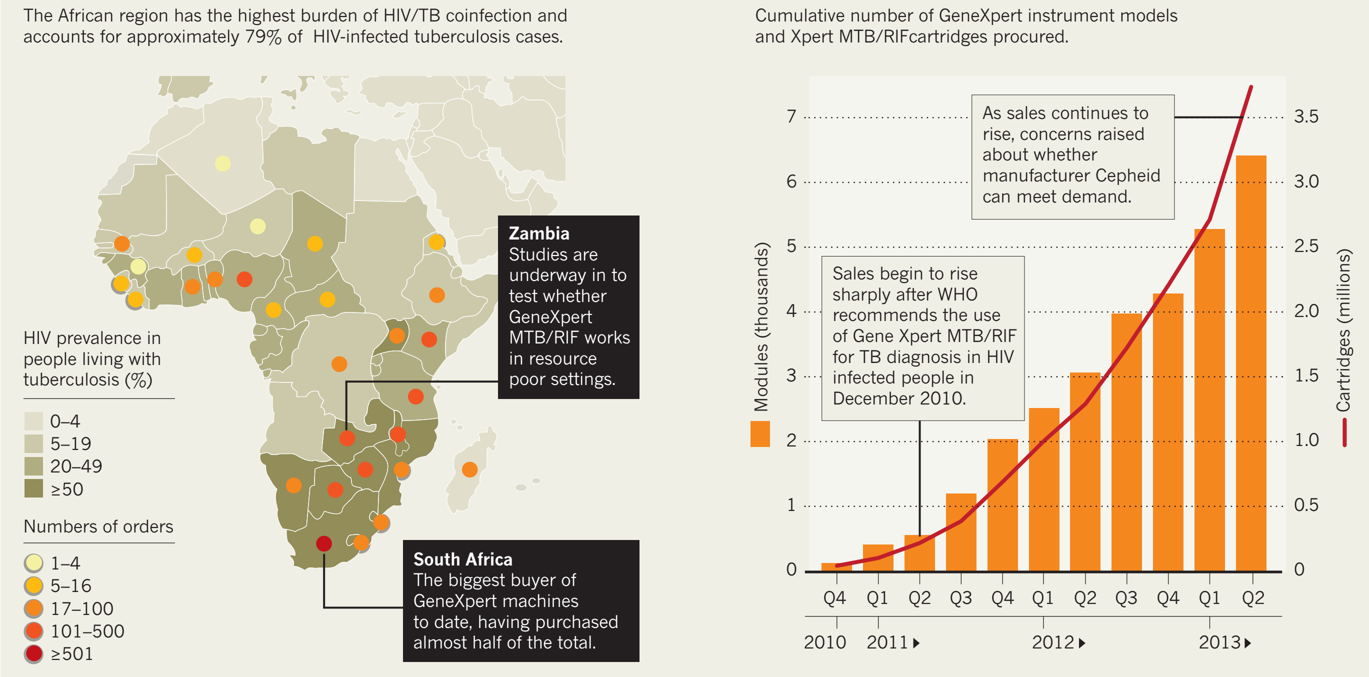
GeneXpert delivers results in as little as two hours, which means patients should be able to wait to receive them. In a busy month, Kanyama clinic will see as many as 500 patients for TB screening. At present, it takes staff between 24 and 48 hours to get results, which means a return trip to the clinic for patients. "Transportation is very difficult," says Kosloff. "If they are told they need to come back tomorrow or next week, they may say, 'Well, I need to decide — is my family going to eat tomorrow, or do I make another trip to the clinic?' If you can't get the results to people the same day, why put this device in such a difficult location?"

There are alternatives to Xpert that are less high-tech. One innovation is an improved form of smear microscopy that uses light-emitting diode (LED) bulbs and fluorescent dyes to show up the bacteria more clearly under the microscope. Although this technique picks up more *M. tuberculosis* than standard microscopy, it also mis-identifies contaminants, leading to false-positive results.

A different approach is to try to rule out, rather than rule in, TB in HIV-infected people

XPERT SALES

Sales of GeneXpert devices, which can help to diagnose tuberculosis in HIV-infected people, have shot up since it was recommended by WHO in 2010.



WHO DATA PROVIDED BY FIND

with suspected infection. Medical diagnostic firm Alere, based in Stockport, UK, has developed a dipstick that tests urine for lipoarabinomannan (LAM), a component of the cell wall of *M. tuberculosis*. Alere's test costs US$3.50 per dipstick and provides results in 25 minutes. It works best in people with a heavily compromised immune system, such as HIV patients with advanced infection. Overall, sensitivity — the proportion of TB it finds — in people also living with HIV is moderate at 40–60%, but its high specificity means that it can rule out TB in almost all uninfected people[3].

Keertan Dheda, a pulmonologist at the University of Cape Town in South Africa, has conducted numerous studies into TB diagnostics. Dheda thinks that, despite its limitations, the dipstick is a valuable tool — particularly given the difficulties of detecting TB in HIV-infected people. "In this category of patients, a urine-orientated approach may be very useful." Chris Smit, director of Alere's infectious disease unit, says he suspects a number of countries are waiting for WHO prequalification of the device before they begin to use it more widely. Alere is working with WHO to get approval by the end of 2014.

BIGGER ISSUES

The GeneXpert system has an additional benefit that these other tests do not: it also checks for resistance to rifampicin, one of the cocktail of four first-line drugs used to treat TB. And if a patient's TB is resistant to rifampicin, there is a good chance it may also be resistant to some of the other drugs. In other words: multidrug-resistant (MDR) TB.

In countries with widespread MDR TB, such as South Africa or throughout Eastern Europe, this ability is valuable. However, in countries where drug resistance remains limited, this resistance test has a low predictive value, meaning that any samples that look like they might be resistant to rifampicin need to be sent to a reference lab to be confirmed. In Zambia, there are only three such labs — and few clinics carry alternative drug regimens.

This is the catch with the use of sophisticated technology in such low-resource settings, says Muyoyeta. "It's a problem because you're identifying these patients and then you can't do anything." This would be especially problematic in more remote parts of the country, which are many days travel from a reference laboratory.

> ***"If they don't invest in TB, then all the money they have put into investing in HIV treatment will just be wiped away."***

Take the case of Tobias Hamooya, back at the Macha Mission Hospital in southern Zambia, where it is more than a day's walk to the nearest small town. This is the second time Hamooya has been diagnosed with TB; the first time he stopped taking the drugs when he started to feel better. Consequently, it is likely that his infection has developed some resistance to first-line drugs, but being so far from a reference laboratory, this would take many weeks to confirm, even if it was picked up by GeneXpert.

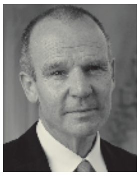
The jury then is still out on whether GeneXpert will live up to its expectations in Zambia. Muyoyeta and her team will publish the findings of the ZAMBART studies in the next few months. She acknowledges the amount of work required before GeneExpert can fulfil its potential. "We refer to GeneXpert as a newborn baby — it needs a lot of attention and is very demanding."

And as Carol Nyirenda, a Zambian activist who sits on the Stop TB Partnership's New Diagnostics Working Group, points out, failing to tackle TB undermines the tremendous progress that has been made to keep people living with HIV alive for longer. "In the earlier days if you were told you have HIV it was like a death sentence," says Nyirenda, who is living with HIV and is a TB survivor. "I've been on medication [ARVs] for close to ten years. TB could come and take me away within a week or two. If they don't invest in TB, then all the money they have put into investing in HIV treatment will just be wiped away." ■

Catherine de Lange *is Naturejobs editor and travelled to Zambia as a new media journalism fellow with the International Reporting Project.*

1. Padmapriyadarsini, C., Narendran, G. & Swaminathan, S. *Indian J. Med. Res.* **134,** 850–865 (2011).
2. World Health Organization. *WHO Monitoring of Xpert MTB/RIF Roll-out* (World Health Organization, 2013).
3. Lawn, S. D. *et al. Lancet* **12,** 201–209 (2012).

PERSPECTIVE




Weigh all TB risks

A narrow definition of risk is hampering the search for new methods of tuberculosis control, say **Christopher Dye** and **Mario Raviglione**.

What factors put people at risk of illness, disability and death? The answers to this question have far-reaching implications: identifying a risk factor suggests interventions that could avoid or alleviate sickness and suffering. Unfortunately, for tuberculosis (TB) and other diseases, current risk assessments are not up to the job.

The Global Burden of Disease Study (GBD) includes an ambitious attempt to pinpoint the major causes of illness worldwide, and to use them to set a global agenda for preventive health care. But, despite listing 67 risk factors in 10 categories[1], the GBD is selective and has little relevance to some important diseases, including TB. A broad set of both disease determinants and factors that limit disease control and treatment should be included in future studies, with this combined set used as the basis for developing a wider range of options for disease prevention and care. This proposal goes beyond TB: for many causes of ill health, an unidentified risk is a missed opportunity.

BROADER CONCEPTS

Most of the risk factors selected by the GBD are environmental exposures, harmful behaviours, such as alcohol abuse, or physiological abnormalities, such as hypertension or high cholesterol. Only 3 of the 67 factors listed are linked to TB — tobacco smoke, alcohol abuse and diabetes. And yet there are clearly many other factors that determine who becomes ill or infectious. For example, migration, urbanization and the way people interact through contact networks are important for the transmission of infection. Genetic factors are also excluded, even though these risks might one day be managed or treated. But perhaps the most significant drawback of the GBD and similar risk assessments is that they do not consider the limitations of current interventions as avoidable risks.

THIS PROPOSAL GOES BEYOND TB: FOR MANY CAUSES OF ILL HEALTH, AN UNIDENTIFIED RISK IS A MISSED OPPORTUNITY.

In TB control, these limitations include poor awareness of symptoms, lack of access to diagnostic and treatment facilities, the prohibitive cost of drugs to treat multidrug-resistant strains, medical malpractice, poor quality of care from health workers, broken drug supply chains, and patients not completing their treatment. Although these shortcomings are not conventionally thought of as risk factors, they account for a large proportion of the avertible burden of disease. To choose the best options for disease control, their importance must be considered next to conventional TB risk factors such as overcrowded housing, diabetes, tobacco smoking, HIV co-infection and under-nutrition. We need, in short, to adopt a more comprehensive view of risk.

ADVERSITY INTO OPPORTUNITY

Finding new ways to reduce TB is a global health priority, and expanding the concept of risk will generate more options for control. By comparing the costs and benefits of possible interventions, we can prioritize the best among them. For example, in India we recently found that the increases in TB risk from diabetes, malnutrition and urbanization are modest compared with the expected positive impact of early detection and treatment[2]. In this setting, the next step is to compare the costs and potential benefits of better case detection and treatment strategies with those of interventions to mitigate other risk factors.

To be comprehensive, this work should go beyond evaluating measures targeted specifically at TB to look at those that have wider benefits for public health, such as health insurance schemes. In this way, the healthcare profession will be encouraged to evaluate interventions that could benefit TB but that lie beyond the reach of current disease control programmes.

CHANGING THE AGENDA

The United Nations (UN) Millennium Development Goals (MDGs) will expire in 2015. During the MDG era, the rise in the TB incidence rate has been halted and reversed, but the decline is still only a disappointing 2% per year globally[3]. Effective TB control programmes should be able to reduce incidence by at least 5–10% each year[4] The new UN agenda for international development will probably focus on poverty reduction and sustainable development[5]. Given limited resources, the challenge for TB control is to take a broader view of risk, setting priorities that overcome a diverse array of obstacles and exploit all possible opportunities. These priorities should include better ways to use existing technologies while promoting the most effective new technologies; working closely with the control of non-communicable diseases; and participating in initiatives to improve health that come not only from the health sector, but also from agriculture, education, finance, industry and housing. This demands a big but potentially rewarding programme of data collection, quantitative analysis and modelling — one that enlarges the idea of risk to unify TB treatment and prevention, and places both in the wider context of health and development. ■

The author(s) alone are responsible for the views expressed in this article. The article does not necessarily represent the decisions, policy or views of the WHO. The WHO retains copyright.

Christopher Dye *is Director of Health Information and* **Mario Raviglione** *is Director of the Global Tuberculosis Programme in the HIV/AIDS, Tuberculosis, Malaria & Neglected Tropical Diseases Cluster at the World Health Organization, Geneva, Switzerland.*
dyec@who.int, raviglionem@who.int

1. Lim, S. S. *et al. Lancet* **380,** 2224–2260 (2012).
2. Dye, C. *et al. PLoS ONE* **6,** e21161 (2011).
3. World Health Organization. *Global Tuberculosis Control: WHO Report 2012* (World Health Organization, 2012).
4. Dye, C. *et al. Annu. Rev Public Health* **34,** 271–286 (2013).
5. United Nations. *A New Global Partnership: Eradicate Poverty And Transform Economies Through Sustainable Development* (United Nations, 2013).

WHO/N. OTTO

Hong Kong and other regions in South-East Asia are trying to get to grips with a resurgence in tuberculosis.

LATENCY

A sleeping giant

Most people infected with Mycobacterium tuberculosis *never get the disease, but predicting who will is turning out to be a complex problem.*

BY COURTNEY HUMPHRIES

Among the elderly population in Hong Kong, a spate of tuberculosis (TB) is worrying public-health officials. But these cases are not the result of recent infection; they signal the emergence of disease from an infection picked up decades ago. "Almost all recent cases in the elderly have arisen through reactivation of long-term latent infections rather than recent new infections," says Benjamin Cowling, an infectious disease epidemiologist at the University of Hong Kong's School of Public Health in China.

Improved standards of living and public-health campaigns in most of the industrialized world dramatically reduced rates of TB during the first half of the twentieth century. But, unlike in the West, Hong Kong, Japan and Korea only began to get to grips with TB after the Second World War. So people who were first infected as children back in the 1940s and 50s — and who have shown no sign of TB for over 50 years — are succumbing to the disease as their immune systems weaken. "Given a high prevalence of latent TB in the elderly, and currently no strategies to stop reactivation," Cowling says, "it is difficult to foresee major changes in TB incidence in the coming years."

By current estimates, two billion people are infected with *Mycobacterium tuberculosis* worldwide, but only 10% will develop active disease in their lifetimes. The rest have what are called latent infections: they are not sick and will not spread the infection to others[1]. The risk of developing disease is roughly 5% in the first 18 months, and 5% over the rest of one's life[2].

This lurking threat is motivating researchers to study TB latency. Are there differences in the bacteria or in the host immune response that determine who gets the disease? Should everyone with latent infection be treated, or are there biomarkers or molecular signatures that could distinguish those who will develop disease?

TRICKY CUSTOMER

Latency is not a well-defined biological condition; it describes anyone who tests positive for TB and does not have clinical symptoms, says David Sherman, a biochemist at the Seattle Biomedical Research Institute in Washington. The assumption has been that *M. tuberculosis* remains viable in people with latent infections, not causing disease but maintaining its potential to do so.

This ability to lie dormant may be an evolutionary strategy of the bacterium. Douglas Young, a microbiologist at the MRC National Institute for Medical Research in London, points out that TB first appeared when early modern humans lived in small communities. "If TB was as virulent then as it is now, you would expect those populations to be wiped out," he says. "There's a big advantage to being latent, letting the population reproduce and then infecting the next generation." The microbe may have only acquired more virulent traits as human settlements grew and it was possible to flourish despite proving fatal to some of its hosts.

But the distinction between latent and active disease may not be as clear-cut as once thought. Ongoing studies of patients with latent TB suggest that although some may have completely controlled the infection, others might have undetected subclinical disease. "Latency encompasses a grab bag of individuals at different points in the spectrum of disease," says Clifton Barry, chief of the Tuberculosis Research Section at the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland.

This emerging view of latent infection is supported by research on the pathology of *M. tuberculosis* infection in animals. JoAnne Flynn, a microbiologist at the University of Pittsburgh in Pennsylvania, worked for years to create a

NATURE.COM
For more on transcriptional profiling, see:
go.nature.com/9ivteq

model of latency in mice. But her breakthrough came about a decade ago when she was working with cynomolgus macaque monkeys, and noticed that about half of the TB-infected animals did not go on to develop disease.

Flynn saw a surprising heterogeneity in lung pathology among individual monkeys — and even within the same individual. She has focused her sights on pockets of inflammation called granulomas: collections of mycobacteria, macrophages and other immune cells, sometimes surrounding dead tissue. Granulomas can vary in size, composition and degree of organization, and are thought to represent an attempt to contain the infection[3]. "It turns out latent monkeys can have a number of granulomas, even one or two that look active," she says. The pathology of the lungs is less severe that those with active infections, but the distinction is not absolute.

SUMMING THE PARTS

Using positron-emission tomography (PET) and computed tomography (CT) imaging, Flynn is able to observe individual lesions in the lung over time, and then perform molecular analyses in more depth. "Each granuloma is its own little world," she says, and whereas some of them are able to control the infection, some are not. Far from hiding or going dormant, the bacterium seems to be involved in an active, dynamic battle in which the immune system fights to keep the disease in check across various individual sites. If that is true, says Sherman, "the distinction between latent TB and active TB becomes an arbitrary one".

The notion that latency represents the sum of its parts would make it difficult to find systemic biomarkers of TB progression. "It suggests that looking at the global immune response is unlikely to be able to predict who is going to do well and who is not," Flynn says. However, studying individual lesions could illuminate the factors that allow host and microbe to coexist for long periods of time, and why the balance sometimes fails.

Barry has been using the same imaging method to study lung lesions in people with active TB, and now in two groups in South Africa and South Korea with latent infections. He has already observed lesions in some people with latent infections and the study "definitely supports the idea that latency in people is incredibly heterogeneous". But it is too early to say whether the extent of the lesions predicts whether people get the disease, as it seems to in monkeys.

To figure out why TB lesions can vary so much, researchers are looking at how the heterogeneity within populations of mycobacteria might influence the microenvironment of the granulomas, and their interaction with the host's immune system. Sarah Fortune, a microbiologist at the Harvard School of Public Health in Boston, Massachusetts, has been working with Flynn to develop molecular tools to analyse the variability of bacteria in monkeys with TB.

Fortune and her team have created bacteria engineered with genetic sequence tags that act as molecular barcodes: one sequence that identifies each strain, and a randomized sequence unique to each bacterium. It allows them to track the fate of different strains and of individual bacteria and their progeny. They have also taken advantage of the fact that when TB bacteria die, they remain in granulomas. Researchers can count the genomes of dead and living bacteria in these bacterial graveyards, helping them to understand the course of the infection. Fortune believes that even bacteria that are genetically identical may differ in gene expression. "There's probably a lot of individuality in the bacterial population that hasn't been appreciated," she says.

For public-health efforts to tackle the reactivation of latent infections, the challenge is first to identify, among the vast numbers of people worldwide with latent infections, those most in need of treatment. "The key issue in managing latent TB is diagnosing people who are at risk of progressing to disease," says Barry.

In 2010, Anne O'Garra, an immunologist at the MRC National Institute for Medical Research, reported progress on this front. Her research showed that the transcriptional profiles of people with active versus latent TB differed. Such profiling could lead to better diagnostic tests to distinguish latent from active infection, or a prognostic tool to determine which people with latent infections will go on to get active disease or to track treatment success in diseased people. Robert Wilkinson, an infectious disease specialist at Imperial College London and the University of Cape Town who collaborates with O'Garra's team, says that they are now testing the approach in larger and more complex groups of patients, including people with HIV and other diseases. "The crucial clinical question is to tell [the profile] apart from diseases that might mimic TB," he says, pointing out that it is not yet clear whether this information can be divined from a transcriptional profile.

"It's guaranteed that we can never break the cycle if we only treat the active cases."

TREAT IT RIGHT

Latent infections can be successfully treated, but if doctors treated everyone then ten people would be getting drugs for every one who would have developed disease, says Richard Menzies, director of the Respiratory Division at the Montreal Chest Institute in Canada.

Consequently, Western governments have made recommendations to screen for and treat latent TB based on risk factors. For example, updated guidelines in 2011 for England and Wales recommend screening all recent migrants from countries where TB is widespread, and suggest that treatment should be considered in people with latent infection who are immunocompromised (such as through HIV infection).

In poor countries where TB infection is endemic and resources are limited, that level of treatment is not feasible, nor is it the priority when so many people with active infections are transmitting disease. But even in developed countries, both doctors and patients are reluctant to initiate treatment in the absence of disease. "There's widespread acceptance of drug treatment for conditions that predispose people to illness, such as hypertension or high cholesterol," says Menzies, "but somehow the notion is you don't have to treat latent TB infection," even though the odds of preventing later disease and of experiencing side effects are comparable.

Ensuring patients complete their treatments, a challenge in patients with active TB, is even more problematic for patients who are not even sick: a long medication regimen for a disease they might never get is a tough sell. Menzies believes that until treatment regimens are available that have a negligible risk of side effects, doctors will continue to be reluctant to prescribe treatment and patients unwilling to accept it.

Developing drugs specifically for latent infection is difficult, as at present drug makers are required to show the efficacy of a drug against replicating bacteria or active disease in mice, but Sherman says that there is a lot of interest among researchers in targeting latency. "Many people now believe that the key to shortening therapy for active disease is to find agents that work on latent disease," says Sherman, because bacteria may be living in a variety of conditions and metabolic states that overlap between the two conditions, and these states may make them more or less susceptible to different drugs.

Isoniazid targets bacteria as they replicate, but drugs developed to exploit other bacterial functions could be more effective at treating latent infections. For instance, researchers at the Institute Pasteur Korea in Seongnam-si have found a compound that targets ATP synthesis in *M. tuberculosis*, a process that is slow in bacteria that are not replicating, but which they still need to survive. Such a drug could benefit both active and latent disease[4].

The situation in Hong Kong illustrates how leaving latent infections untreated is a stumbling block to achieving very low incidence rates. "It's guaranteed that we can never break the cycle if we only treat the active cases," says Barry. Cowling agrees. "Latent TB is going to be a problem for many years to come," he says. And he believes that what is happening now in more developed areas such as East Asia is a harbinger of future problems in less developed countries that are still grappling with active disease. ■

Courtney Humphries *is a freelance writer based in Boston, Massachusetts.*

1. Esmail, H., Barry, C. E. III & Wilkinson R. J. *Drug Discov. Today* **17,** 514–521 (2012).
2. Zumla, A. *et al. N. Eng. J. Med.* 368, 745–755 (2013).
3. Ramakrishnan, L. *Nature Rev. Immunol.* **12,** 352–366 (2012).
4. Pethe, K. *et al. Nature Med.* **19,** 1157–1160 (2013).

NEIL WEBB

TRANSMISSION

Control issues

Once tuberculosis takes hold in a population it can be hard to control, but scientists are finding new ways to understand and stop its spread.

BY EWEN CALLAWAY

The USS *Richard E. Byrd* left its Norfolk, Virginia port in January 1965, bound for the Mediterranean. The 4,000-tonne destroyer carried more than 300 sailors — one of whom was infected with *Mycobacterium tuberculosis*, the bacteria that cause tuberculosis (TB). By the time he left the ship more than a year later, coughing violently, almost half the crew had acquired silent infections and seven of his cabin mates and other close contacts had full-blown TB.

"It was just a beautiful case of being able to pin down when transmission happened and when disease happened," says Clifton Barry, head of the Tuberculosis Research Section at the National Institute of Allergy and Infectious Disease in Bethesda, Maryland.

The USS *Byrd* and other locally limited outbreaks offer an opportunity to monitor the spread of the bacterial disease. They provide critical epidemiological data for scientists studying TB transmission in less cramped quarters, and highlight the challenges to combatting a disease that can spread wildly before even being noticed. "We never find the cases before they've infected 10 or 15 other people," Barry says.

Many experts see reducing the spread of *M. tuberculosis* infection as the only way to manage the disease. "Transmission is really the dominant problem we have in TB control," says Chris Dye, an epidemiologist at the World Health Organization (WHO) in Geneva, Switzerland. "We know how to treat patients in clinics and cure them and save their lives and reduce illness, but what we've been far less successful at doing is cutting out transmission."

STEMMING THE TIDE

TB infections come in two general forms: active and latent. People with latent TB, like most of the infected sailors aboard the USS *Byrd*, are not symptomatic and are unlikely to develop the disease (see 'Latency: A sleeping giant', page S14). A long course of antibiotics can rid the body of the low levels of bacteria they carry, but even without treatment, they will probably never pass on their infection to others.

Most TB control efforts target the active form of the disease. Active infections can spread the mycobacterium like wildfire, even before patients develop symptoms such as a cough — and start expelling droplets filled with the infectious bacteria into the air for other people to inhale. Mark Perkins, chief scientific officer at the Foundation for Innovative New Diagnostics, in Geneva, uses the analogy of mopping up "the spill" of highly infectious people with active TB "before you turn off a tap that is barely dribbling".

Poor quality medical care can exacerbate transmission. In countries such as India it is not uncommon for patients with active TB to visit four or more healthcare providers over a period of up to several months before getting a correct diagnosis[1], and unregulated medical practices are rife. "Patients are spending far too long in the community transmitting TB to other people, before they end up in a clinic that can diagnose their TB and put them on the right treatment," says Dye.

In most parts of the world, a diagnosis of active TB is confirmed by observing the tiny bacilli in a sample of sputum put under the microscope (see 'Diagnosis: Waiting for results', page S10). But by the time *M. tuberculosis* reaches high enough numbers to be seen in this way, patients are already infectious. "You've missed the boat on stopping transmission," says Perkins.

Efforts to control TB should focus on lowering the costs of diagnostics (and drugs) for governments, healthcare providers and, ultimately, patients, says Dye. "The TB control community are not by themselves going to remedy the problem of poor health systems, but they're going to have to make a contribution."

BEYOND BORDERS

Among the countries with a high burden of TB in Eastern Europe, Africa and Asia, one stands out as having had particular success at getting to grips with TB transmission. Estonia, a nation of 1.3 million people perched on the Baltic Sea, experienced a surge in TB in the early 1990s, after the collapse of the Soviet Union — and with it the healthcare infrastructure it provided.

With support from nearby Scandinavian countries, the Estonian government implemented a TB control programme, involving medical training, public-health screening and treatment programmes, says Manfred Danilovits, a physician at Tartu University Hospital who heads the programme. As a result, incidence dropped dramatically — from over 50 cases per 100,000 in 2002 to about 25 cases in 2011 (ref. 2). The TB control programme is now experimenting with ways to tackle drug-resistant TB and co-infection with HIV, says Danilovits. For instance, some people living with HIV, many of whom will have acquired the infection through intravenous drug use, who are also infected with TB can get all their medicines — antibiotics, antiretrovirals and, if necessary, methadone — at one clinic instead of three, improving their adherence to a strict treatment programme.

Understanding transmission is critical to stemming the rise of TB in Western countries as well. London earned a reputation as the 'TB capital of Europe' after cases rose by nearly 50% between 1999 and 2009, from 2,309 to 3,450 (ref. 3). More than half of TB cases in regions such as the United Kingdom are in immigrants from high-burden countries, says Iacopo Baussano, an epidemiologist at WHO's International Agency for Research on Cancer in Lyon, France, who has studied the effectiveness of TB screening among new immigrants.

In a 2012 survey[4] of 29 high-income countries, Baussano and colleagues found that 25 of them screen immigrants for active TB infections — typically with a skin test or chest X-ray, or by checking for symptoms such as a cough. Only a third of those countries screen people seeking entry before arrival; nearly all of them screen migrants soon after they arrive. Just 16 of the countries screened any immigrants for latent TB infection.

"Transmission is really the dominant problem we have in TB control."

Undocumented immigrants and refugees are at the greatest risk of TB infection, Baussano says, because they often emigrate from countries with a high incidence of TB and live in high-density housing that fosters transmission.

Existing TB screening programmes do a good job of identifying cases in new immigrants, with infection rates typically matching those of the immigrants' country of origin. But it has proven more difficult to prevent new TB cases in people who have recently migrated to cities such as London. His work suggests that marginalized immigrant populations should be screened more closely for TB in the years after they arrive.

TB risk tends to decrease the longer individuals live in a country, presumably after people gain access to health care and more stable housing. And Baussano says that there is no evidence to suggest that the TB circulating among some groups of immigrants ever spreads into the general community.

Unfortunately, in the European Union, there is a political impediment to coordinated TB screening. Procedures vary widely between countries and even between cities, and many cases slip through the region's open borders. "It's not easy to integrate the systems," says Baussano.

MARIANNE LOORENTS/OHTULEHT

TB patients await treatments in Estonia.

TAKING THE STRAIN

One major challenge to understanding the transmission of different forms of TB, such as those that are resistant to drugs or spread more quickly, is identifying them, says Ruth McNerney, a molecular microbiologist at the London School of Hygiene and Tropical Medicine. Related TB strains can differ very little at the genetic level, and most laboratory tests cannot tell them apart. McNerney's team is turning to sequencing the entire genome (made up of approximately 4 million building blocks) of clinical isolates of TB, to track their spread more closely.

McNerney and colleague Taane Clarke are building a reference library of TB varieties so that researchers can easily trace a strain they have identified back to those circulating elsewhere in the world. Even with the cost of genome sequencing falling and states such as the United Kingdom integrating the technology into routine health care, genome sequencing is unlikely to influence care in poor, high-burden countries any time soon, McNerney concedes. But she believes that data gained from sequencing will be used to develop simpler, cheaper tests to discern different strains of TB and help inform treatment.

TACKLING RESISTANCE

Sequencing is also helping researchers to analyse outbreaks of drug-resistant TB, McNerney says. Her team recently looked at strains of drug-resistant TB from people in Uganda who had been hospitalized twice with the disease, comparing the *M. tuberculosis* genomes of the two infections. In some patients, the strains differed very little, suggesting that drug resistance had resulted from poor compliance with antibiotics. But in other patients they differed markedly, says McNerney, and the second infection often matched the genome of a strain circulating in the same clinic. "There seems to be problem of multidrug resistance being shared amongst re-treatment cases," she says.

Whole-genome sequencing can also help reveal the underlying social causes of a TB outbreak. A team led by Jennifer Gardy, a molecular epidemiologist at the British Columbia Centre for Disease Control in Vancouver, investigated an outbreak in Vancouver that resulted in 41 cases of active TB between 2006 and 2008. Genome sequencing revealed that the epidemic was really two outbreaks, caused by distinct strains of TB that had emerged at the same time[5]. Gardy's findings suggested that the rise of crack-cocaine use in the city sparked the outbreaks, with crack houses becoming centres of TB transmission.

Her team also found that a handful of 'superspreaders' were responsible for transmitting most of the TB cases. Biological factors, such as the levels of TB bacilli in a person's sputum or cough, or a suppressed immune system that allows rampant bacterial replication, may explain this phenomenon, she says. But Gardy's work indicates that social contact is often the most important element of an outbreak. "TB fundamentally is a social disease," she says. "Social factors are what makes an outbreak, versus one sick person who doesn't transmit to anybody else." ■

Ewen Callaway *is a Senior News Reporter for* Nature.

1. Dye, C. *Indian J. Med. Res.* **135,** 737–744 (2012).
2. The World Bank. *Incidence of Tuberculosis* (Per 100,000 People) (The World Bank, 2013).
3. Zumla, A. *Lancet* **377,** 10–11 (2011).
4. Pareek, M. *et al. Emerg. Infect. Dis.* **18,** 1422–1429 (2012).
5. Gardy, J. L. *et al. N. Engl. J. Med.* **364,** 730–739 (2011).